Design and Optimization of Wireless Networks for Large Populations

Alonso Ariel Silva Allende

N° D'ORDRE: 2010-01-TH

THÈSE DE DOCTORAT

SPECIALITÉ : PHYSIQUE

École Doctorale "Sciences et Technologies de l'Information, des Télécommunications et des Systèmes"

Présentée par :

Alonso Ariel SILVA ALLENDE

Sujet :

Planification et optimisation des réseaux sans fil pour des grandes populations

Soutenue le 07 Juin 2010 devant les membres du jury :

M. Eitan Altman	INRIA Sophia Antipolis-Méditerranée	Directeur de Thèse
M. Tamer Başar	University of Illinois at Urbana-Champaign	Rapporteur
M. Pierre Bernhard	INRIA Sophia Antipolis-Méditerranée	Examinateur
M. Mérouane Debbah	École Supérieure d'Électricité (Supélec)	Co-directeur de Thèse
M. Philippe Jacquet	INRIA Paris-Rocquencourt	Examinateur
M. Marc Mézard	Université Paris-Sud	Examinateur
M. Hikmet Sari	École Supérieure d'Électricité (Supélec)	Examinateur
M. Stavros Toumpis	Athens University of Economics and Business	Rapporteur

"Un ejército de invisibles manos ha labrado la tierra,
ha levantado tu casa, ha servido tu mesa, para que tú puedas aprender.
Ahora, esas innumerables manos - las más desposeídas - se tienden hacia ti
con el gesto de la necesidad. Te piden simplemente lo que les pertenece."

Cristián del Campo, Doctor en Economía, chileno.

Résumé

La croissance explosive des réseaux sans fil et l'augmentation du nombre de dispositifs sans fil ont soulevé de nombreuses difficultés techniques dans la planification et l'analyse de ces réseaux. Nous utilisons la modélisation continue, utile pour la phase initiale de déploiement et l'analyse à grande échelle des études régionales du réseau. Nous étudions le problème de routage dans les réseaux ad hoc, nous définissons deux principes d'optimisation du réseau: le problème de l'utilisateur et du système. Nous montrons que les conditions d'optimalité d'un problème d'optimisation construit d'une manière appropriée coïncide avec le principe de l'optimisation de l'utilisateur. Pour fonctions de coût différentes, nous résolvons le problème de routage pour les antennes directionnelles et omnidirectionnelles. Nous trouvons également une caractérisation des voies du coût minimum par l'utilisation extensive du Théorème de Green dans le cas d'antennes directionnelles. Dans de nombreux cas, la solution se caractérise par une équation aux dérivées partielles. Nous proposons l'analyse numérique par éléments finis qui donne les limites de la variation de la solution en ce qui concerne les données. Lorsque nous permetons la mobilité des origines et destinations, on trouve la quantité optimale de relais actif. Dans les réseaux MIMO et canaux de diffusion MIMO, nous montrons que, même lorsque la chaîne offre un nombre infini de degrés de liberté, la capacité est limitée par le rapport entre la taille du réseau d'antennes la station de base et la position des mobiles et la longueur d'onde du signal. Nous constatons également l'association optimale mobile pour différentes politiques et distributions des utilisateurs.

Abstract

The growing number of wireless devices and wireless systems present many challenges on the design and operation of these networks. In this thesis, we focus on massively dense ad hoc networks and cellular systems. We use the continuum modeling approach, useful for the initial phase of deployment and to analyze broad-scale regional studies of the network. We study the routing problem in massively dense ad hoc networks, and similar to the work of Nash [1], and Wardrop [2], we define two principles of network optimization: user- and system-optimization. We show that the optimality conditions of an appropriately constructed optimization problem coincides with the user-optimization principle. For different cost functions, we solve the routing problem for directional and omnidirectional antennas. We also find a characterization of the minimum cost paths by extensive use of Green's theorem in directional antennas. In many cases, the solution is characterized by a partial differential equation. We propose its numerical analysis by finite elements method which gives bounds in the variation of the solution with respect to the data. When we allow mobility of the origin and destination nodes, we find the optimal quantity of active relay nodes.

In Network MIMO systems and MIMO broadcast channels, we show that, even when the channel offers an infinite number of degrees of freedom, the capacity is limited by the ratio between the size of the antenna array at the base station and the mobile terminals position and the wavelength of the signal. We also find the optimal mobile association for the user- and system-optimization problem under different policies and distributions of the users.

Contents

Résumé en Français

Introduction

La croissance explosive des réseaux sans fil et l'augmentation du nombre de dispositifs sans fil tels les téléphones 3G, les ordinateurs portables WiFi, et les réseaux de capteurs sans-fil, ont soulevé de nombreuses difficultés techniques dans la planification et l'analyse des réseaux sans fil.

Les réseaux sans fil peuvent être essentiellement classifiés en deux catégories : réseaux d'accès sans fil et les réseaux *ad hoc* sans fil.

- Les réseaux d'accès sans fil offrent habituellement la connectivité au réseau filaire via l'environnement sans fil. Les réseaux d'accès sans fil comprennent, par exemple, les systèmes cellulaires comme 2G GSM, 3G UMTS, etc., et les systèmes WiFi comme 802.11 WLANs.
- Les réseaux *ad hoc* sans fil sont des réseaux décentralisés qui peuvent être installés et désinstallés de manière dynamique. Les réseaux de capteurs sans fil, VANETs, etc., sont des exemples de réseaux *ad hoc* sans fil.

Dans cette thèse, nous nous sommes intéressés aux réseaux sans fil massivement denses dans les systèmes cellulaires et les réseaux *ad hoc* sans fil massivement denses. Le terme "massivement dense" sera formellement défini par la suite. Cette thèse a débuté sur l'idée d'un projet de recherche sur la planification et l'analyse des réseaux sans fil en établissant un parallèle avec les dits "outils physiques" au sens large. Au début de ce travail, nous avons découvert qu'il existait une grande quantité de travaux préalables et en cours de développement dans une autre communauté scientifique: la communauté des réseaux de transport. Nous allons introduire brièvement l'objet de la recherche dans cette communauté et les liens avec notre travail sur la modélisation de réseaux sans fil. Dans la littérature de la communauté de réseaux de transport, la modélisation des problèmes de circulation et d'équilibre pour un réseau de transport est classifiée en modélisation discrète et modélisation continue :

- Dans la modélisation discrète, la demande d'accès est supposée concentrée dans des dits barycentres; ces barycentres sont des points hypothétiques qui représentent une zone de couverture, chaque liaison routière entre les zones au sein du réseau étant alors des liaisons entre les barycentres. Cette approche de modélisation est couramment adoptée pour la planification et l'analyse détaillée de réseaux de transport.

- La modélisation continue est utilisée pour la phase initiale de planification et la modélisation à large échelle des études régionales. L'objectif de cette approche est d'obtenir la tendance générale et le schéma de la distribution et du choix de voyage des utilisateurs. L'objectif est aussi l'étude des modifications de ces deux facteurs en réponse aux changements de politique dans le système de transport à l'échelle macroscopique, plutôt qu'une description détaillé du réseau. L'hypothèse fondamentale est que la différence dans les caractéristiques de modélisation, comme le coût du voyage et la demande entre zones adjacentes au sein d'un réseau, est relativement faible par rapport à la variation de l'ensemble du réseau. Par conséquent, les caractéristiques d'un réseau, tels que l'intensité de flux, la demande, et le coût du voyage, peuvent être représentés par des fonctions mathématiques régulières.

La modélisation continue a plusieurs avantages par rapport à l'approche discrète. Premièrement, elle réduit la taille du problème dans les réseaux, car la taille du problème dans la modélisation continue dépend de la méthode qui est utilisée pour rapprocher la région de modélisation mais ne dépend pas du réseau lui-même. Ainsi, une méthode d'approximation efficace, telle que la méthode des éléments finis (FEM en anglais), peut largement réduire la taille du problème. Cette réduction du problème permet de gagner du temps de calcul et de la mémoire. Deuxièmement, moins de données sont nécessaires dans la configuration du modèle dans une modélisation continue. La modélisation continue peut être caractérisée par un petit nombre de variables spatiales, peut être miss en place avec une quantité beaucoup plus restreinte de données que la modélisation discrète, qui elle a besoin de données pour tous les liens. Cela rend la modélisation continue plus adaptée à l'étude macroscopique que la modélisation discrète. Dans la phase initiale de conception du réseau, la collecte des données prend du temps et requiert de complexes calculs, de sorte que les ressources nécessaires pour l'entreprendre ne sont généralement pas disponibles. Il n'y a ainsi généralement pas de données suffisantes dans le système pour mettre en place un modèle détaillé. Enfin, la modélisation continue nous donne une meilleure compréhension des caractéristiques globales d'un réseau. De manière assez surprenante, les avancées de cette communauté ne sont pas très répandues dans la communauté des télécommunications. Dans ce contexte, nous avons d'abord analysé le problème de routage dans les réseaux *ad hoc* sans fil de capteurs qui doivent transporter des paquets à travers le réseau. Les parallèles avec le transport routier ont été abordées précédemment dans le contexte d'allocation optimale des ressources grâce à la programmation linéaire, par Hitchcock [3] et Kantorovich [4] (qui a plus tard partagé le prix Nobel avec Koopmans) ainsi que par Koopmans [5] et Dantzig [6]. Dans ces modèles, toutefois, la congestion liée au réseau de transport n'était pas considérée. En 1952, Wardrop [2] a établi deux principes de transport dans l'utilisation du réseau, qui sont appelés respectivement l'optimisation de l'utilisateur et l'optimisation du système: le premier principe déclare que les voyageurs vont choisir l'itinéraire de voyage d'origine en destination de manière indépendante. Dans une situation d'équilibre, le temps de parcours des routes utilisés entre une paire origine-destination est invariante. De plus, ce temps est inférieur à ce qu'obtiendrait un seul véhicule, s'il prend une voie libre. Le second principe reflète la situation dans laquelle un contrôleur central achemine les flux de circulation de manière optimale entre origines et destinations de manière à minimiser le coût total du réseau. En 1956, Beckmann, McGuire, et Winsten [7] ont été les premiers à fournir une formulation mathématique rigoureuse des conditions énoncées par le premier principe de Wardrop, qui a permis de trouver la solution du problème d'équilibre de la circulation dans le réseau pour des fonctions de coût croissantes par rapport au flux sur les liens. En particulier, il a été démontré que les conditions d'optimalité sous la forme de conditions de Karush-Kuhn-Tucker [8, 9] d'une programmation mathématique appropriée conduisent à un problème d'optimisation qui coïncidé avec l'affirmation selon laquelle les coûts du voyage sur les routes utilisées (chemins reliant chaque paire origine-destination dans un réseau de transport) ont les même coûts de voyage et ces coûts de voyage sont minimum. Par conséquent, aucun voyageur, agissant de façon unilatérale n'aura de motivation à modifier sa trajectoire (en supposant un comportement rationnel) donnée si son coût du voyage (délai de voyage) est minimum. Ainsi, un problème dans lequel il existe de nombreux décideurs agissant de manière indépendante, et comme plus tard a été également noté par Dafermos et Sparrow [10] concurrentes dans le sens de Nash [1], pourraient être reformulé (sous des hypothèses appropriées qui seront définies ultérieurement) comme un problème d'optimisation convexe avec une fonction objectif unique, soumise à des contraintes linéaires et de non-négativité des hypothèses du flux sur le réseau.

Beckmann [11] a noté la pertinence des concepts d'équilibre dans les réseaux de télécommunications. Dans une autre étude, Bertsekas et Gallager [12] ont observé des similitudes entre les réseaux de communication et les réseaux de transport. Les travaux sur le paradoxe de Braess [13], par la suite, ont fourni l'un des principaux liens entre les réseaux du transport et les réseaux en informatique. En 1990, Cohen et Kelly [14] ont décrit un paradoxe analogue à celui de Braess dans le cas d'un réseau de files d'attente. Ce paradoxe continue à être étudié dans le contexte du trafic routier [15, 16], ainsi que dans la communauté de réseaux de télécommunications [17, 18, 19].

Dans la communauté des réseaux de télécommunications, les réseaux massivement denses ont apporté de nombreuses difficultés. Quand un réseau a un nombre croissant de nœuds, la modélisation et l'analyse du réseau est beaucoup plus difficile et parfois impossible à résoudre. Quand nous parlons de réseaux denses, mous supposons une forte séparation entre le niveau macroscopique, correspondant à des distances typiques entre les sources et ses destinations, et le niveau microscopique, correspondant à des distances entre les nœuds voisins. Lorsque le système est suffisamment grand, le modèle macroscopique nous donnera une meilleure description de ce réseau et l'on peut tirer des conclusions de ses propriétés à partir de considérations microscopiques. Dans la modélisation continue, la description détaillée de la solution optimisée est sacrifiée mais le modèle macroscopique permet de préserver une quantité d'informations suffisantes afin de donner une meilleure description du réseau et la dérivation des résultats intéressants dans des configurations différentes.

Les méthodes inspirées de la physique ont été utilisées pour l'étude des réseaux *ad hoc* massivement denses avec les travaux initiaux de Jacquet [20], et de Kalantari et Shayman [21, 22]. Dans ce contexte, un certain nombre de groupes de recherche ont travaillé sur les réseaux *ad hoc* massivement denses en utilisant des outils issus de la recherche en optique géométrique [20] ainsi que de l'électrostatique (voir par exemple [23, 24, 25], et l'étude [26] sur ce sujet, ainsi que les références contenues dans ces textes). Nous allons les décrire dans les sections suivantes en accordant une attention particulière dans le Chapitre 1, Partie II. Les paradigmes physiques ont permis de minimiser divers critères liés au problème de routage. Hyytia et Virtamo ont proposé dans [27] une approche basée sur l'équilibrage des charges de flux dans le réseau en faisant valoir que, si le plus court chemin (ou chemin de coût minimal) est suivi, certaines parties du réseau devraient transporter plus de trafic que les autres et donc consommer plus d'énergie que les autres. Cela aboutirait à une vie plus courte du réseau étant donné que quelques parties auront épuisé leur énergie plus tôt que d'autres, et plus tôt que dans un réseau avec une charge de flux équilibrée.

Le développement de la théorie originelle de routage dans les réseaux massivement denses parmi la communauté de réseaux *ad hoc* a émergé de manière complètement indépendante de la théorie existante de routage dans les réseaux massivement denses élaborée au sein de la communauté des réseaux de transport. L'approche adoptée en 1952 par Wardrop [2] et par Beckmann [28] est encore un domaine de recherche active dans cette communauté, voir [29, 30, 31, 32, 33] et références dans celles-ci.

Les principales contributions de la thèse seront le déploiement optimal des nœuds de relais dans le cas des réseaux *ad hoc* sans fil massivement denses et le déploiement optimal des stations de base et d'association des mobiles dans le cadre de systèmes cellulaires. Nous associons dans ces thèses différentes approches pour résoudre ces problématiques, comme par exemple, la théorie du contrôle et de la théorie du transport optimal, et fournissons de nouvelles méthodologies pour le problème de routage.

Réseaux *ad hoc* sans fil

Les problèmes de cheminement

Le problème du chemin le plus court ou chemin de coût minimal consiste à trouver un chemin entre deux sommets (ou nœuds) tel que la somme des poids des arêtes traversées par ce chemin est minimisé. L'objective est ainsi de trouver un chemin des origines jusqu'aux destinations de telle façon que le coût total du trajet (considéré comme la somme des coûts de transmission de toutes les liens entre les nœuds qui appartiennent au trajet) est minimale parmi tous les chemins reliant les origines aux destinations. Les algorithmes les plus importants pour faire face à ce problème sont :

- L'algorithme de Dijkstra qui résout le problème de cheminement en présence d'une seule source et d'une seule destination, en utilisant la famille de notations de Landau, à un temps d'exécution O(nombre d'arêtes^2).

- L'algorithme de Ford-Bellman qui résout le problème d'acheminement pour une seule source lorsque les poids des arêtes peuvent être négatifs, et qui a un temps d'exécution O(nombre de sommets x nombre d'arêtes).

- L'algorithme de Floyd-Warshall connu aussi sous le nom de l'algorithme WFI ou Roy-Floyd, qui résout toutes les paires de chemins les plus courts, a un temps d'exécution O(nombre de arêtes^3)

Pour une explication sur ces algorithmes, voir par exemple [34]. Un problème important pour tous ces algorithmes est qu'ils ont besoin d'un temps d'exécution exponentielle dans le nombre d'entrées et que la quantité de données requises par la modélisation discrète peut ne pas être disponible. L'autre problème avec la résolution du problème du chemin le plus court a été mentioné par Hyytiä et Virtamo dans [27]. Comme mentionné précédemment, ils proposent une approche basée sur l'équilibrage des charges de flux dans le réseau en faisant valoir que, si le plus court chemin (ou de minimisation des coûts) est suivi, certaines parties du réseau devraient transporter plus de trafic que les autres et donc consommer plus d'énergie que les autres. Cela aboutirait à une vie plus courte du réseau donné étant donné qu'un sous-partie du réseau aura épuisé son énergie plus tôt que l'ensemble du réseau, et plus tôt que dans un réseau avec une charge de flux équilibrée.

Fonctions de coût

Dans l'optimisation d'un protocole d'acheminement dans les réseaux *ad hoc*, ou d'optimisation du positionnement des nœuds, l'un des points de départ est la détermination de la fonction de coût qui reflète le coût de transporter d'un paquet à travers le réseau. Pour déterminer celle-ci, une spécification du réseau est nécessaire, qui comprend les éléments suivants :

- Une topologie du réseau
- Une règle de transfert que les nœuds vont utiliser pour sélectionner le prochain nœud pour la transmission d'un paquet.
- Les coûts de transmission pour transmettre un paquet entre nœuds intermédiaires.

Fonctions de coûts indépendants de la congestion

Une métrique utilisée souvent dans l'Internet pour déterminer les coûts du chemin entre une origine et une destination est le nombre de transmissions nécessaires pour transmettre un paquet entre cette origine et cette destination. Dans le cadre de réseaux *ad hoc*, le nombre de transmissions entre nœuds intermédiaires est proportionnel au délai prévu dans le chemin, si les délais des files d'attente sont négligeables par rapport aux délais de transmission sur chaque paire origine-destination. Ce critère est insensible aux interférences ou à la congestion du réseau. Nous supposons qu'il ne dépend que du rayon de transmission.

Fonction de coût dépendant de la congestion

Une autre fonction de coût plus générale est de considérer que le coût peut dépendre d'une mesure de la congestion dans le réseau. Pour mesurer la congestion, nous considerons la

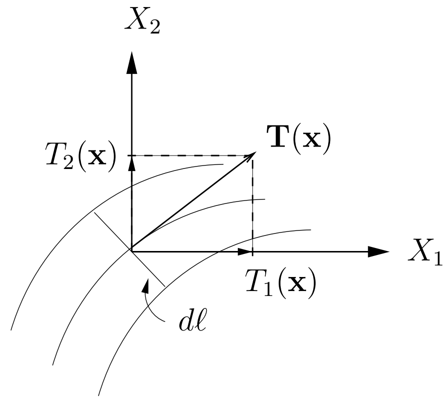


Figure 1: Flux de l'information $\mathbf{T}(\mathbf{x})$ a travers le segment de ligne incremental $d\ell$, decomposé dans sa composante horizontal $T_1(\mathbf{x})$ (dans la direction X_1) et sa composante vertical $T_2(\mathbf{x})$ (dans la direction X_2).

fonction de flux de l'information $\mathbf{T}(\mathbf{x})$, mesurée en bps/m, de tel façon que ses directions coincident avec la direction du flux de l'information dans le point $\mathbf{x}$, et $\|\mathbf{T}(\mathbf{x})\|$ est le taux avec lequel le flux de l'information croisse un segment linéaire perpendicular à $\mathbf{T}(\mathbf{x})$ centré en $\mathbf{x}$, c.à.d, $\|\mathbf{T}(\mathbf{x})\|\,\varepsilon$ nous donne la quantité total de traffic que croisse un segment linéaire de longueur infinitesimal ε, centré dans le point $\mathbf{x}$, et perpendiculaire à $\mathbf{T}(\mathbf{x})$.

Nous considerons maintenant qu'une fonction de coût, dénotée c, peut dépendre du flux de trafic, dénoté $\mathbf{T}$, qui passe par un point particulier $\mathbf{x}$. Dans ce cas, nous supposons que la fonction de coût dépend de la quantité du flux de l'information qui passe par ce point mais elle ne dépendra pas de la direction de ce flux. En plus, elle peut dépendre aussi de l'endroit où la transmission a lieu. Dans ce cas, $c = c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)$.

Modélization de l'équation de conservation

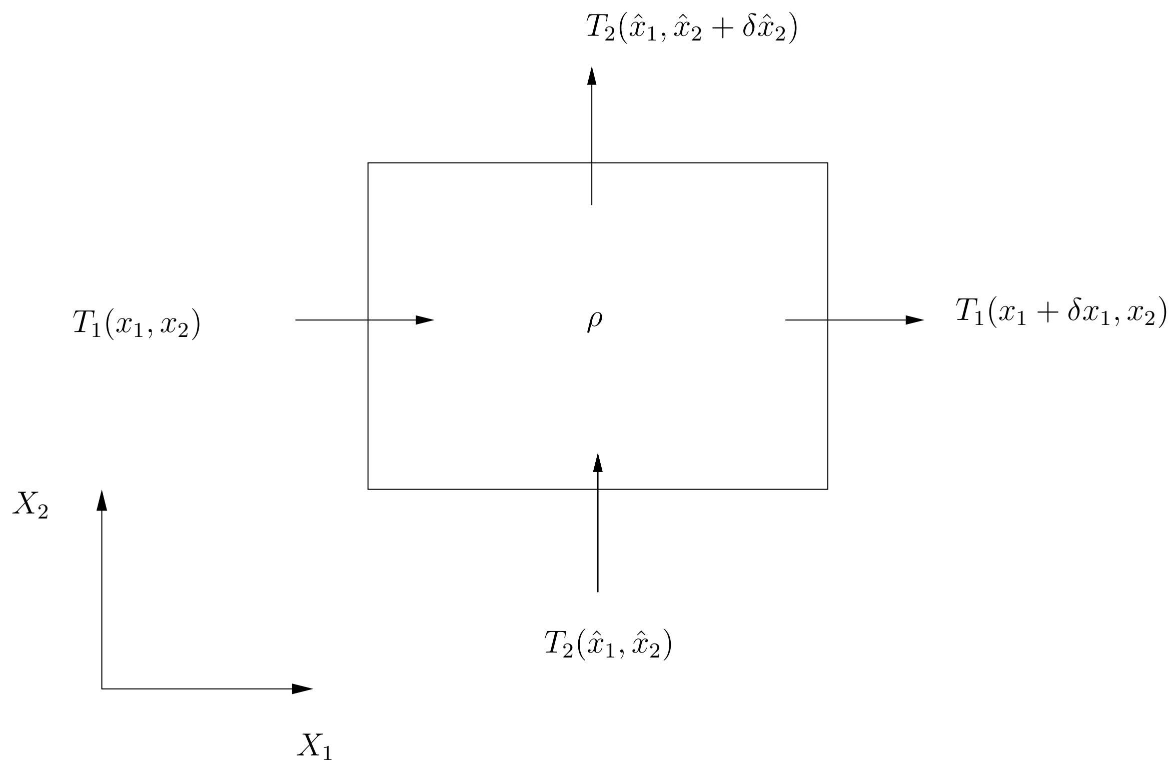
Nous considèrons un réseau dans le plan à deux dimensions $X_1 \times X_2$. Nous considèrons aussi, la fonction d'information continue $\rho(\mathbf{x})$, mesurée en bps/m^2, telle que dans les points $\mathbf{x}$ où $\rho(\mathbf{x}) > 0$, l'information est créée par les sources, de telle façon que le taux auquel l'information est créée dans un zone de taille infinitésimale dA_ϵ, centré à l' endroit $\mathbf{x}$, est $\rho(\mathbf{x})dA_\epsilon$. De la même manière, aux points où $\rho(\mathbf{x}) < 0$, il existe des centres d'agrégation où il existe une récupération de l'information, telles que le taux auquel l'information est reçue en un point de taille infinitésimale dA_ϵ, centré en $\mathbf{x}$, est égale à $-\rho(\mathbf{x})dA_\epsilon$.

Étant donné que cette situation se déroule en tout point du domaine, il s'ensuit que nécessairement:

$$\nabla \cdot \mathbf{T}(\mathbf{x}) := \frac{\partial T_1(\mathbf{x})}{\partial x_1} + \frac{\partial T_2(\mathbf{x})}{\partial x_2} = \rho(\mathbf{x}),$$

où "$\nabla\cdot$" est l'opérateur de divergence.

Notons que la dernière équation est la version différentielle du théorème de Gauss (aussi appelé théorème de Gauss-Ostrogradsky ou théorème de Green).



Antennes directionnelles

Optimisation d'utilisateurs et fonctions de coûts indépendants de la congestion

Nous autorisons le coût de transmission local c_1 pour une transmission horizontale (dans la direction de l'axe X_1) à être différent de celui du transmission local c_2 pour une transmission verticale (dans la direction de l'axe X_2). On suppose que les coûts de transmission local, c_1 et c_2, ne dépendent pas du flux de l'information $\mathbf{T}$. Le coût de transmission d'un paquet à travers d'un chemin p est donnée par l'intégrale

$$\mathbf{c}_p = \int_p \mathbf{c} \cdot \, d\mathbf{x}.$$

Soit $V(\mathbf{x})$ le coût minimum pour aller d'un point $\mathbf{x}$ a un ensemble B. Par conséquence, on obtient que pour tout $\mathbf{x}$ dans l'ensemble des destinations $V(\mathbf{x}) = 0$. Nous obtenons récursivement

$$V(\mathbf{x}) = \min\left(c_1(\mathbf{x})\, dx_1 + V(x_1 + \, dx_1, x_2), c_2(\mathbf{x})\, dx_2 + V(x_1, x_2 + \, dx_2)\right). \tag{1}$$

Cela peut être écrit comme une équation de Hamilton-Jacobi-Bellman:

$$\forall \mathbf{x} \in \mathcal{D}, \quad 0 = \min\left(c_1(\mathbf{x}) + \frac{\partial V(\mathbf{x})}{\partial x_1}, c_2(\mathbf{x}) + \frac{\partial V(\mathbf{x})}{\partial x_2}\right); \qquad \forall \mathbf{x} \in B\,,\ V(\mathbf{x}) = 0\,. \tag{2}$$

Si la fonction V est dérivable, alors (dans des conditions appropriées) elle est la solution unique de (2). Dans le cas où V n'est pas partout différentiable (dans des conditions appropriées) elle est la solution unique de l'équation de viscosité (2) (voir par exemple [35, 36]).

Il existe des nombreuses méthodes numériques pour résoudre l'équation de Hamilton-Jacobi-Bellman (HJB). Par exemple, on peut discrétiser l'équation de HJB et trouver un problème de programmation dynamique discrète pour lequel il existe des méthodes qui nous donnent des solutions d'une manière efficace. Si l'on répète ces mesures de discrétisation diverses, alors on sait que la solution du problème discret converge vers la solution de viscosité du problème original (dans des conditions appropriées) quand la taille de l'étape suivante par rapport a l'étape actuelle, tend vers zéro [35].

Caractérisation des chemins de coût minimum

Nous considérons maintenant notre modèle d'antennes directionnelles dans une zone rectangulaire dénotée R, définie par la courbe fermée $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$. Nous étudions le cas où les transmissions peuvent aller de Nord en Sud ou d'Ouest en Est. Cette convention est prise en suivant la notation utilisée par Dafermos dans [29]. Grâce à la modélisation continue les chemins optimaux définis comme les chemins qui permettent d'atteindre le coût de transmission minimal de paquets sont obtenus. Nous allons étudier deux problèmes :

- Chemin optimal d'une seule source à une seule destination: nous cherchons le chemin de coût de transmission minimal de la route entre deux nœuds: une seule source et une seule destination.
- Chemin optimal d'une seule source à plusieurs destinations: nous cherchons le chemin de coût de transmission minimal entre une seule source qui peut choisir entre un ensemble de destinations celui qui minimise le coût de transmission de la route.

Nous allons utiliser pour cela le théorème suivant pour la caractérisation des chemins optimaux:

Theorem 0.0.1 (Théorème de Green) *Soit $\mathcal{S}$, une courbe plane simple, positivement orienté et $\mathcal{C}^1$ par morceaux, $\mathcal{D}$ le domaine compact lisse du plan délimité par S et $P\,dx+Q\,dy$ une 1-forme différentielle sur $\mathbb{R}^2$. Si P et Q ont des dérivées partielles continues sur une région ouverte incluant D, alors :*

$$\int_{\mathcal{S}} P\,dx + Q\,dy = \int_{\mathcal{D}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\,dy.$$

Considérons la fonction

$$U(\mathbf{x}) = \frac{\partial c_2}{\partial x_1}(\mathbf{x}) - \frac{\partial c_1}{\partial x_2}(\mathbf{x}),$$

alors la structure du chemin du coût minimum va dépendre du signe de cette fonction U.

En fait, on prouve le théorème suivant:

Theorem 0.0.2 (Chemin optimal d'une seule source à une seule destination) *Si on suppose qu'un point d'origine $\mathbf{x}^o = (x_1^o, x_2^o)$ veut envoyer un paquet a une point de destination $\mathbf{x}^d = (x_1^d, x_2^d)$ et les deux points sont à l'intérieur du rectangle R.*

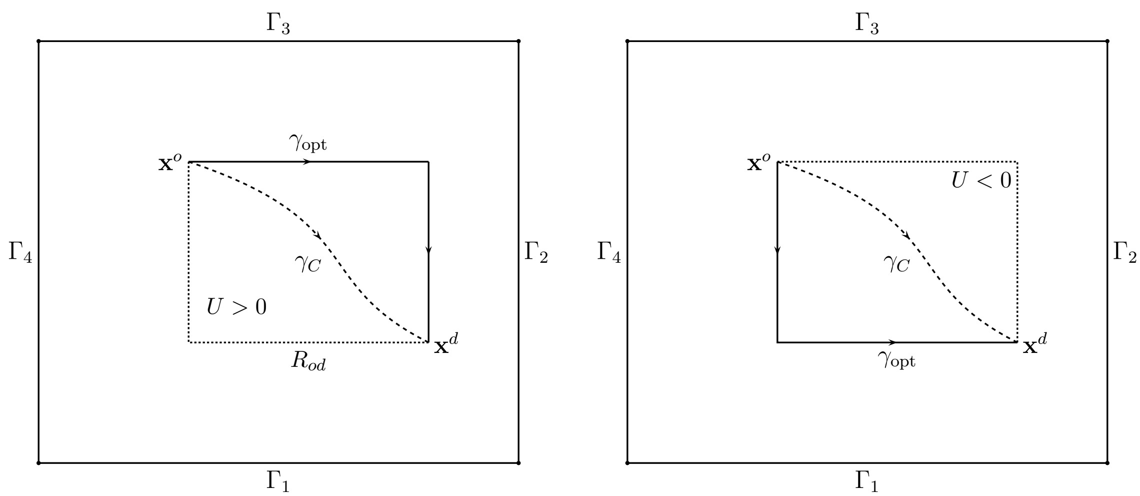
1. *Si la fonction U est "presque partout" (p.p.) positive dans l'intérieur du rectangle R_{od}, défini par les deux points, le chemin optimal γ_{opt} est l'union d'une ligne horizontale γ_H et d'une ligne verticale γ_V (voir Fig. 2(a)). Plus précisément: $\gamma_{\text{opt}} = \gamma_H \cup \gamma_V$ où*

$$\gamma_H = \{(x_1, x_2) \text{ such that } x_1^o \leq x_1 \leq x_1^d, x_2 = x_2^o\},$$
$$\gamma_V = \{(x_1, x_2) \text{ such that } x_1 = x_1^d, x_2^o \leq x_2 \leq x_2^d\}.$$

2. *Si la fonction U est "presque partout" positive dans l'intérieur du rectangle R_{od}, défini par les deux points, le chemin optimal γ_{opt} est l'union d'un ligne horizontale γ_H et d'une ligne verticale γ_V (voir Fig. 2(b)). Plus précisément, $\gamma_{\text{opt}} = \gamma_V \cup \gamma_H$ où*

$$\gamma_V = \{(x_1, x_2) \text{ such that } x_1 = x_1^o, x_2^o \leq x_2 \leq x_2^d\},$$
$$\gamma_H = \{(x_1, x_2) \text{ such that } x_1^o \leq x_1 \leq x_1^d, x_2 = x_2^d\}.$$

3. *Dans les deux cas précedents, γ^{opt} est unique presque sûrement.*



(a) Case $U > 0$. (b) Case $U < 0$.

Figure 2: Chemins Optimaux (a) quand $U > 0$ p.p. et (b) quand $U < 0$ p.p. dans l'intérieur du rectangle défini par le point d'origine $\mathbf{x}^o$ et le point de destination $\mathbf{x}^d$.

Theorem 0.0.3 *(Chemin optimal d'une seule source à plusieurs destinations)*
Si on suppose qu'un point d'origine $\mathbf{x}^o$ veut envoyer un paquet vers un point de la frontière $\Gamma_1 \cup \Gamma_2$

1. *Si la fonction U est "presque partout" positive dans l'intérieur du rectangle R, et le coût dans la frontière Γ_1 est non-positive et dans la frontière Γ_2 est non-négative, alors le chemin optimal γ_{opt} est la ligne horizontale γ_H (voir Fig 3)*

2. *Si la fonction U est "presque partout" négative dans l'intérieur du rectangle R, et le coût dans la frontière Γ_1 est non-négative et dans la frontière Γ_2 est non-positive, alors le chemin optimal γ_{opt} est la ligne vertical γ_V (voir Fig. 4)*

3. *Dans les deux cas précédents, γ^{opt} est unique presque sûrement.*

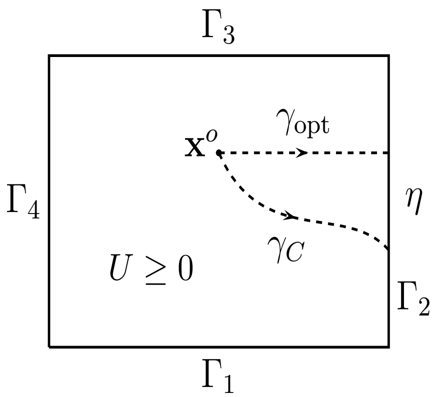


Figure 3: Chemin optimal quand $U > 0$ p.p.

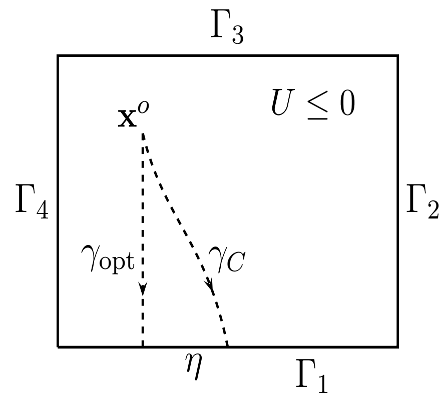


Figure 4: Chemin optimal quand $U < 0$ p.p.

Coût dépendent de la congestion

Comme avant, nous autorisons le coût de transmission local c_1 (dans la direction de l'axe X_1) à être différent de celui du transmission local c_2 (dans la direction de l'axe X_2). Cette fois-ci nous permetons aussi que le coût de transmission c_1 puisse dépendre du flux de trafic T_1 (dans la direction de l'axe X_1). De la même façon, le coût de transmission c_2 peut dépendre du flux de trafic T_2 (dans la direction de l'axe X_2).

Soit $V(\mathbf{x})$ le coût minimal pour aller d'une source $\mathbf{x}$ à la frontière B dans une situation d'équilibre. L'équation (1) est toujours valide mais cette fois-ci avec c_1 et c_2 qui dépendent du flux de l'information T_1 et T_2, respectivement. Donc l'équation (2) devient

$$\forall \mathbf{x} \in \mathcal{D}, \quad 0 = \min_{i=1,2} \left(c_i(\mathbf{x}, T_i) + \frac{\partial V(\mathbf{x})}{\partial x_i} \right), \quad \forall \mathbf{x} \in B \,, V(\mathbf{x}) = 0 \,. \tag{3}$$

Notez que cette méthode peut être considérée comme une généralisation de la méthode d'optimisation connus dans la programmation dynamique dans la modélisation discrète. En particulier, la dernière équation est une généralisation de "l'équation de Bellman" également connu comme "l'équation de programmation dynamique".

Nous notons que, si $T > 0$, alors par la définition d'équilibre, le minimum est atteint dans (3). Donc (3) implique les relations suivantes pour $i = 1, 2$:

$$c_i(\mathbf{x}, T_i) + \frac{\partial V^k}{\partial x_i} = 0 \quad \text{if} \quad T_i^k > 0, \tag{4a}$$

$$c_i(\mathbf{x}, T_i) + \frac{\partial V^k}{\partial x_i} \geq 0 \quad \text{if} \quad T_i^k = 0. \tag{4b}$$

Il s'agit d'un ensemble d'équations aux dérivées partielles couplées difficile d'analyser sans faire plus d'hypothèses.

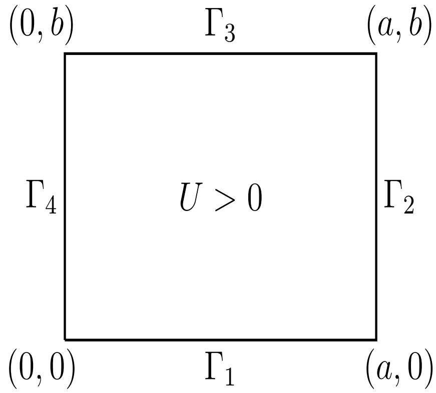


Figure 5: Le rectangle R défini par les frontières $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$, lorsque $U > 0$.

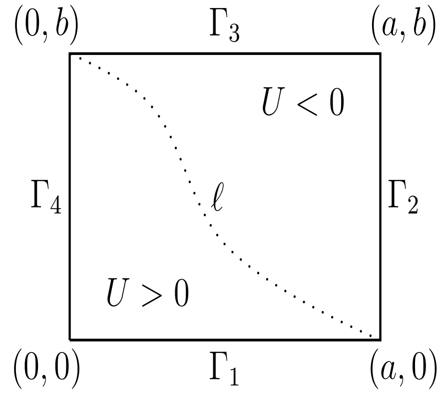


Figure 6: Le cas de deux régions séparées par une courbe.

Transformation de Beckmann

De manière similaire à Beckmann et al. [7] pour la modélisation discrète, nous avons transformé le problème du chemin de coût minimal en un système équivalent de minimisation. Nous allons analyser maintenant le problème d'une seule classe. Pour ceci, nous constatons que les équations (4a)-(4b) ont exactement la même forme que les conditions de Karush-Kuhn-Tucker ([8])-([9]), sauf que les fonctions de coût $c_1(\mathbf{x}, T_1)$ et $c_2(\mathbf{x}, T_2)$ dans la première équation sont remplacées par $\partial g(\mathbf{x}, \mathbf{T})/\partial T_1(\mathbf{x})$ et $\partial g(\mathbf{x}, \mathbf{T})/\partial T_2(\mathbf{x})$ respectivement, dans le second. Nous allons introduire une fonction de potentielle ψ, définie par

$$\psi(\mathbf{x}, \mathbf{T}) = \sum_{i=1,2} \int_0^{T_i} c_i(\mathbf{x}, s) \mathrm{d}s$$

Alors pour chaque $i \in \{1, 2\}$

$$c_i(\mathbf{x}, T_i) = \frac{\partial \psi(\mathbf{x}, \mathbf{T})}{\partial T_i}.$$

Par conséquent, le flux de l'information dans une situation d'équilibre est celui obtenu à partir du système de problèmes d'optimisation où nous utilisons ψ comme le coût local. Notre conclusion est comme suit:

Theorem 0.0.4 *Soit x^* une solution pour le problème d'optimisation suivant :*

$$\min_{T(\cdot)} \int_\Omega \psi(\mathbf{x}, \mathbf{T})\, d\mathbf{x}$$

soumis aux conditions suivantes:

$$\nabla \cdot \mathbf{T}(\mathbf{x}) = \rho(\mathbf{x}), \qquad \forall \mathbf{x} \in \Omega.$$

Alors, c'est un équilibre de Wardrop.

Les messages doivent aller d'une région des sources ou origins de l'information $\mathcal{O}$ à une région disjointe des récepteurs de l'information $\mathcal{R}$ (dans les réseaux de capteurs sans fil, il correspondrait aux centres d'agrégation de données). Ces deux régions sont supposées d'être situés à des portions disjointes de la frontière. L'intensité $\sigma(x_1, x_2)$ de la production des messages sur $\mathcal{O}$ est donné, tandis que l'intensité $\rho(x_1, x_2)$ de reception des messages sur $\mathcal{R}$ est inconnue. C'est seulement supposer que ceux-ci sont cohérentes: le débit total de messages émis et reçus sont égaux. Sur le reste de la frontière (denotés par $\mathcal{F}$), aucun message devrait entrer ni sortir de Ω, *i.e.* il s'agit d'un région interdit à croisser.

Soit $\mathcal{Q} = \mathcal{O} \cup \mathcal{F}$ et on extende la fonction σ dans tout $\mathcal{Q}$ par $\sigma(\mathbf{x}) = 0$ en $\mathcal{F}$. Nous modelons les conditions de la frontière comme:

$$\forall \mathbf{x} \in \mathcal{Q}, \quad \langle \mathbf{n}(\mathbf{x}), \mathbf{T}(\mathbf{x}) \rangle = -\sigma(\mathbf{x}). \tag{5}$$

Nous supposons seulement dans cette partie qu'il n'y a pas des sources ni destinations des messages dans l'interior de Ω. Ce qu'on modelise comme la contrainte:

$$\forall \mathbf{x} \in \Omega, \quad \nabla \cdot \mathbf{T}(\mathbf{x}) = 0\,. \tag{6}$$

Le coût de congestion par paquet transmis $c(x_1, x_2, \phi)$ (par exemple en termes de retards, ou utilisation de l'énergie) à chaque point de Ω est une fonction du point et de l'intensité ϕ du flux de messages à travers ce point.

Nous voulons étudier la politique optimal de routage et sa relation avec l'équilibre de Wardrop ou l'optimalité de l'utisateur.

Le coût de la congestion par paquet c est censé être une fonction $\mathcal{C}^1$ strictemente positive $c(\mathbf{x}, \phi) : \Omega \times \mathbb{R}_+ \to \mathbb{R}_+$, croissante et convexe dans ϕ pour chaque $\mathbf{x}$. Le coût total de la congestion sera considéré comme

$$G(\mathbf{T}(\cdot)) = \int_\Omega c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)) \cdot \|\mathbf{T}(\mathbf{x}))\| \; d\mathbf{x}. \tag{7}$$

Le chemin suivi par un paquet est spécifié par sa direction de voyage $e_\theta = (\cos\theta, \sin\theta)$ au long de son parcours, en fonction de $\dot{\mathbf{x}} = e_\theta$. Le coût par un paquet qui voyage de $\mathbf{x}_0 \in \mathcal{O}$ au moment t_0 en $\mathbf{x}_1 \in \mathcal{R}$ au moment t_1 est

$$J(e_\theta(\cdot)) = \int_{\mathbf{x}_0}^{\mathbf{x}_1} c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|) \sqrt{\mathrm{d}x^2 + dy^2} = \int_{t_0}^{t_1} c(\mathbf{x}(t), \|\mathbf{T}(\mathbf{x}(t))\|) \; dt\,. \tag{8}$$

Notons ici que le "temps" t peut être un temps fictif, lié au temps physique, disons τ, par $d\tau = c\; dt$, par exemple. Par consequence, c est l'inverse de la vitesse du voyage, un retard dû à la congestion, et J est le temps pris par le message pour aller de la source à la destination.

Optimisation du système

Le cas différentiable

Soit $C(\mathbf{x}, \phi) := c(\mathbf{x}, \phi)\phi$. C'est une fonction convexe en ϕ et coercive, c.à.d., elle tend vers l'infini avec ϕ. Par conséquent, $\mathbf{T}(\cdot) \mapsto G(\mathbf{T}(\cdot))$ est continue, convexe et coercive. De plus, les contraintes sont linéaires. Donc, un optimum existe, et nous pouvons appliquer le Théorème de Karush-Kuhn-Tucker.

Nécessairement

$$\forall \mathbf{x} : f^\star(\mathbf{x}) \neq 0\,, \quad \mathrm{D}_2 C(\mathbf{x}, \|f^\star(\mathbf{x})\|) \frac{f^\star(\mathbf{x})}{\|f^\star(\mathbf{x})\|} = \nabla p(\mathbf{x})\,. \tag{9}$$

Il résulte de cette équation que $p(\cdot) \in H^1(\Omega)$, et aussi que la première intégrale dans le côté droit doit être zéro pour chaque G en $V_{\mathcal{Q}}$. En choisissant maintenant $g \in V_{\mathcal{Q}}$, il s'ensuit que

$$p(\cdot) \in H^1_{\mathcal{R}}. \tag{10}$$

Manque de dérivabilité

En remplaçant cela dans le sous-différentiel de $\mathcal{L}$, nous obtenons, pour $\mathbf{x} \in \Omega_0$,

$$\exists q(\mathbf{x}) \text{ such that } \|q(\mathbf{x})\| \leq \mathrm{D}_2 C(\mathbf{x}, 0) \text{ and } \forall g \in V_{\mathcal{Q}}\,, \int_{\Omega_0} (q(\mathbf{x}) - \nabla p(\mathbf{x})) g(\mathbf{x})\ d\mathbf{x} = 0\,.$$

En combinant les deux cas, nous arrivons à la conclusion que, pour qu'une fonction $f^\star(\cdot) \in V$ avec un ensemble nul Ω_0 soit optimale, il doit exister un $p(\cdot) \in H^1_{\mathcal{R}}$ de tel façon que

$$\begin{array}{ll} \forall \mathbf{x} \in \Omega\,, & \|\nabla p(\mathbf{x})\| \leq \mathrm{D}_2 C(\mathbf{x}, 0)\,, \\ \forall \mathbf{x} \in \Omega - \Omega_0\,, & \nabla p(\mathbf{x}) = \mathrm{D}_2 C(\mathbf{x}, \|\mathbf{T}(\mathbf{x})^\star\|) \frac{1}{\|f^\star(\mathbf{x})\|} f^\star(\mathbf{x})\,. \end{array} \tag{11}$$

Nous pouvons remarquer que la première condition ci-dessus implique aussi

$$\forall \mathbf{x} : f^\star(\mathbf{x}) \neq 0\,, \quad \|\nabla p(\mathbf{x})\| = D_2 C(\mathbf{x}, \|f^\star(\mathbf{x})\|).$$

Dans l'ensemble, le problème de la détermination de l'optimum $f^\star$ est équivalente (si ce système a une solution unique) à déterminer simultanément $f^\star$ et p en satisfaisant (5),(6) et (11).

Ce système a certainement au moins une solution, puisque notre problème est convexe, coercive avec des contraintes affines, et a donc un minimum. Unicité, d'autre part, est loin d'être simple. On peut remarquer que l'on pourrait chercher à retrouver les deux fonctions scalaires phi et p, satisfaisant

$$\begin{array}{ll} \forall \mathbf{x} : \phi(\mathbf{x}) \neq 0\,, & \|\nabla p(\mathbf{x})\| = \mathrm{D}_2 C(\mathbf{x}, \phi(\mathbf{x}))\,, \\ \forall \mathbf{x} : \phi(\mathbf{x}) = 0\,, & \|\nabla p(\mathbf{x})\| \leq \mathrm{D}_2 C(\mathbf{x}, 0)\,, \\ \forall \mathbf{x} \in \mathcal{R}\,, & p(\mathbf{x}) = 0\,, \end{array}$$

et qui imposent en plus les contraintes (5) et (6)

$$f^\star(x) = \frac{\phi(\mathbf{x})}{\mathrm{D}_2 C(\mathbf{x}, \phi(\mathbf{x}))} \nabla p(\mathbf{x})\,.$$

Optimisation de l'utilisateur (équilibre de Wardrop)

Supposons que le flux de messages obéit aux conditions nécessaires obtenues ci-dessus. Nous voulons étudier s'il est optimal pour un seul message en suivant la route prescrite par $f^\star$, *i.e.*, d'une intégrale de ligne de ce champ, en supposant que son seul écart de ce schema n'aurait pas d'effet sur l'ensemble de la congestion (c'est ce qu'on appelle hypothèse d'"atomicité").

Nous étudions l'optimisation de la fonction d'objectif (8) via son équation de Hamilton-Jacobi-Bellman. Soit $V(\mathbf{x})$ la fonction de retour, donc elle doit être la solution de viscosité de

$$\begin{array}{ll} \forall \mathbf{x} \in \Omega\,, & \min_\theta \langle e_\theta, \nabla V(\mathbf{x})\rangle + c(\mathbf{x}, \|f^\star(\mathbf{x})\|) = 0\,, \\ \forall \mathbf{x} \in \mathcal{R}\,, & V(\mathbf{x}) = 0\,. \end{array}$$

alors

$$\begin{array}{ll} \forall \mathbf{x} \in \Omega\,, & -\|\nabla V(\mathbf{x})\| + c(\mathbf{x}, \|f^\star(\mathbf{x})\|) = 0\,, \\ \forall \mathbf{x} \in \mathcal{R}\,, & V(\mathbf{x}) = 0\,. \end{array} \tag{12}$$

Et la direction optimale de voyage est contraire par $\nabla V(\mathbf{x})$, *i.e.*, $e_\theta = -\nabla V(\mathbf{x})/\|\nabla V(\mathbf{x})\|$.

Notons que ce système d'équations et similaire à celui précedent en remplaąnt $p(\mathbf{x})$ par $-V(\mathbf{x})$, et $\mathrm{D}_2 C(\mathbf{x}, \phi)$ par $c(\mathbf{x}, \phi)$. Nous arrivons à la conclusion que l'équilibre de Wardrop peut être obtenu en résolvant le problème d'optimisation du système dans lequel la fonction de coût est remplacé par $\int_0^\phi c(\mathbf{x}, s) ds$. Celle-ci est la version continue de la fonction de potentiel de Beckman *et al.* [7]. Cette transformation a été fréquement utilisée dans le contexte de trafic routier mais seulement par une fonction de coût particulière [37, 38, 33, 39]. Cette équivalence a été montrée dans [37, 38].

Coût du Monôme

Dans le cas où $c(\mathbf{x}, \phi) = c(\mathbf{x})\phi^\alpha$, alors $C(\mathbf{x}, \phi) = \alpha c(\mathbf{x}, \phi)$, et par conséquent, les deux systèmes déquations coïncident dans le domaine $\{\mathbf{x} \mid f^\star(\mathbf{x}) \neq 0\}$. Nous allons montrer que, pour $\phi(\cdot)$ donné, p est défini de façon unique. Nous avons donc la propriété suivante:

Proposition 0.0.1 *Pour un coût du monôme, tout équilibre dans le plan où $\Omega_0 = \emptyset$ est un équilibre de Wardrop.*

Coût linear de la congestion

Nous étudions ici le cas typique simple, où le coût de congestion est linéaire:

$$c(\mathbf{x}, \phi) = \frac{1}{2} c(\mathbf{x})\phi, \text{ et alors } C(\mathbf{x}, \phi) = \frac{1}{2} c(\mathbf{x})\phi^2\,.$$

Ensuite, $\mathcal{L}$ est dérivable partout, et la condition nécessaire d' optimalité est simplement qu'il doit exister $p : \Omega \to \mathbb{R}^2$ tel que $\nabla p(\mathbf{x}) = c(\mathbf{x}) f^\star(\mathbf{x})$.

En remplaçant ceci dans (5) et (6), nous nous retrouvons avec une équation elliptique simple avec les conditions de frontière mixtes Dirichlet - Neuman:

$$\left.\begin{array}{ll} \forall \mathbf{x} \in \Omega\,, & \nabla(\frac{1}{c(\mathbf{x})}\nabla p(\mathbf{x})) = 0\,, \\ \forall \mathbf{x} \in \mathcal{Q}\,, & \dfrac{\partial p}{\partial n}(\mathbf{x}) = c(\mathbf{x})\sigma(\mathbf{x})\,, \\ \forall \mathbf{x} \in \mathcal{R}\,, & p(\mathbf{x}) = 0\,, \end{array}\right\} \tag{13}$$

pour laquelle on obtient facilement l'existence et l'unicité de la solution.

Dans le but de trouver une solution numérique de notre problème, nous considérons la Méthode des Éléments Finis (MEF), qui est très utilisée dans la modélisation numérique des systèmes physiques dans plusieurs domaines comme: l'Électromagnétisme, la Dynamique des Fluides, etc.

La formulation variationnelle d'un problème consiste en général à chercher $u \in V$ tel que

$$(\mathrm{VP}) \left\{ \begin{array}{l} a(u,v) = l(v) \\ \forall\, v \in V. \end{array} \right.$$

En ce sens, si nous considérons les fonctions $a(\cdot,\cdot)$ et $l(\cdot)$ définies ci dessous

$$a(u,v) = \int_\Omega \frac{1}{c} \nabla u \cdot \nabla v \, dx \quad \text{and} \quad l(v) = \int_\mathcal{Q} \sigma \cdot v \, dx.$$

dans l'espace $V = H^1_\mathcal{R}(\Omega)$, notre problème se réduit à sa formulation variationnelle, dont la solution sera la fonction p.

Dans notre cas, la fonction bilinéaire $a(\cdot,\cdot)$ est V-elliptique, symétrique et continue dans $H^1(\Omega)$ et la fonction linéaire $l(\cdot)$ est bornée. Donc nous pouvons utiliser le théorème de Lions-Lax-Milgram et conclure qu'il existe une solution unique.

Ce théorème nous fournit non seulement l'existence et l'unicité de la solution mais aussi l'information par rapport à la stabilité de cette solution quand les données initiales du problème changent. Ceci veut dire que la solution dépend continument des données initiales.

Dans les réseaux mobiles *ad hoc*, la fonction de la densité d'information ρ, la fonction du flot de trafic $\mathbf{T}$ et la fonction de densité des nœuds η, définies auparavant, peuvent varier dans le temps, *i.e.*, $\rho = \rho(\mathbf{x}, t)$, $\mathbf{T} = \mathbf{T}(\mathbf{x}, t)$, et $\eta = \eta(\mathbf{x}, t)$. Nous considérons le problème de routage pour un intervalle de temps défini $t \in [t_i, t_f)$ où t_i représente l'instant initial et t_f est l'instant final.

Réseaux Cellulaires

Nous considérons le réseau $\mathcal{D}$ avec un grand nombre des terminaux mobiles repartis avec une distribution intégrable de $\lambda(x,y)$ qu'on peut dimensionner de telle façon que

$$\iint_\Omega \lambda(x,y)\, dx\, dy = 1.$$

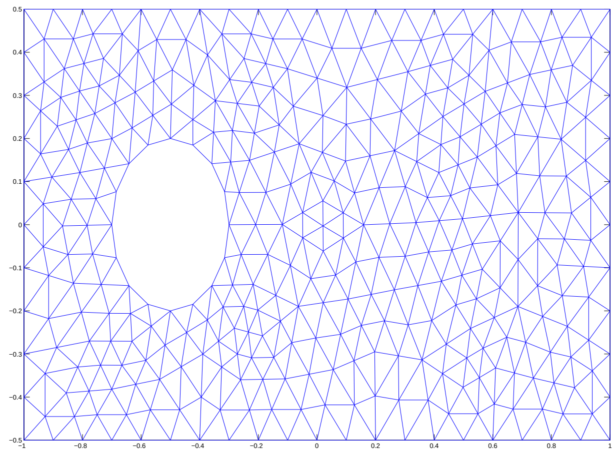


Figure 7: Triangulation du domain $[-1,1]\times[-0.5,0.5]\backslash D((-0.5,0),0.2)$.

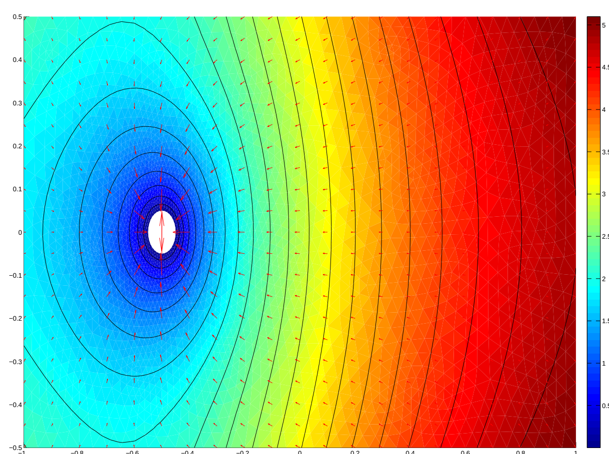


Figure 8: Solution pour un réseau des capteurs sans fils dans le domain $[-1,1]\times[-0.5,0.5]\backslash D((-0.5,0),0.2)$.

Alors, le nombre d'utilisateurs dans un espace A est

$$N\left(\iint_A \lambda(x,y)\right)$$

où N est le nombre totale de terminaux mobiles.

Nous supposons que dans le réseau il y a K stations de base $\mathrm{BS}_1, \mathrm{BS}_2, \ldots, \mathrm{BS}_K$ situées dans les positions $(x_1, y_1), (x_2, y_2) \ldots, (x_K, y_K)$.

Pour la voie montante (transmission des terminaux mobiles aux stations de base) nous considérons comme mesure le SINR (rapport signal et interférence plus bruit). Pour la voie descendante (transmission des stations de base aux terminals mobils) entre stations de base voisins, ils transmet dans des canaux orthogonaux (comme en OFDMA), et l'interférence entre stations de base distantes est négligeable, donc au lieu de considérer le SINR comme mesure, on utilise le SNR (rapport signal à bruit).

Theorem 0.0.5 *Considérons le problème* (P1)

$$\operatorname*{Min}_{C_i} \sum_{i=1}^{K} \iint_{C_i} \left[F(d_i(x,y)) + s_i\left(\iint_{C_i} \lambda(\omega,z)\, d\omega\, dz\right)\right] \lambda(x,y)\, dx\ dy,$$

où C_i est la partition des cellules de Ω. Supposons que s_i sont des fonctions continument différentiable, croissante, et convexe. Le problème (P1) *admet une solution que vérifie*

$$(\mathrm{S1}) \begin{cases} C_i = & \{x : F(d_i(x,y)) + s_i(N_i) + N_i \cdot s_i'(N_i) \leq \\ & \leq F(d_j(x,y)) + s_j(N_j) + N_j \cdot s_j'(N_j)\} \\ N_i = & \iint_{C_i} \lambda(\omega,z)\, d\omega\, dz. \end{cases}$$

Theorem 0.0.6 *Nous considérons le problème* (P2)

$$\operatorname*{Min}_{C_i} \sum_{i=1}^{K} \iint_{C_i} \left[F(d_i(x,y)) \cdot m_i\left(\iint_{C_i} \lambda(\omega,z)\, d\omega\, dz\right)\right] \lambda(x,y)\, dx\, dy$$

où C_i est une partition en cellules du réseau Ω. Nous supposons que les fonctions $m_i, i \in \{1, \ldots, K\}$ sont dérivables. Le problème (P2) *admet une solution que vérifie*

$$(\mathrm{S2}) \begin{cases} C_i = & \{x\ :\ m_i(N_i)F(d_i(x,y))\, \lambda(x,y) + U_i(x,y) \leq \\ & \leq m_j(N_j)F(d_j(x,y))\, \lambda(x,y) + U_j(x,y)\} \\ U_i = & m_i'(N_i) \iint_{C_i} F(d_i(x,y))\lambda(x,y)\, dx\, dy \\ N_i = & \iint_{C_i} \lambda(\omega,z)\, d\omega\, dz. \end{cases}$$

Nous supposons qu'on veut minimiser la fonction de puissance total du réseau en assurant un certain débit moyen de θ à chaque mobile dans le système en utilisant la politique de round robin donnée par le problème (RR)

$$\operatorname*{Min}_{C_i} \sum_{i=1}^{K} \iint_{C_i} \sigma^2 (R^2 + d_i(x,y)^2)^{\xi/2} (2^{N_i\theta} - 1) \lambda(x,y)\, dx\, dy.$$

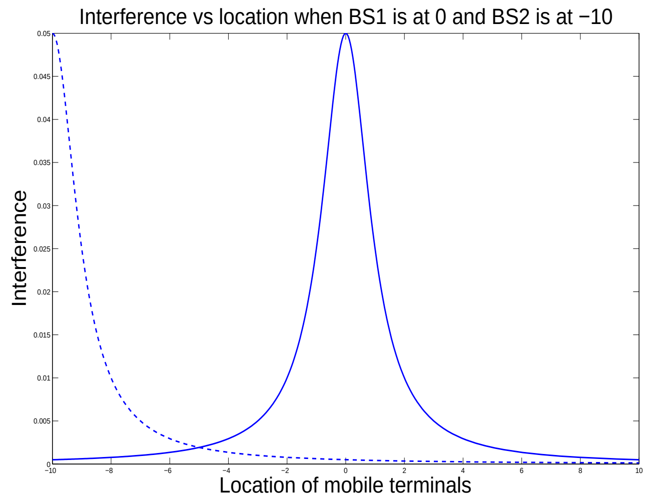


Figure 9: Interference comme fonction de l'ubication de terminal mobiles quand BS_1 est dans la position 0 (ligne plaine) et BS_2 dans -10 (ligne en pointillés).

Nous notons que ce problème est un problème de transport optimal comme (P1) avec une fonction de coût donnée par

$$F(d_i(x,y)) = \sigma^2(R^2 + d_i(x,y)^2)^{\xi/2}$$

$$m_i(x,y) = (2^{N_i\theta} - 1)$$

Synthèse et conclusions

Dans ce manuscrit, nous avons étudié la planification et l'analyse des réseaux sans fil massivement denses. Nous nous sommes intéressés à l'approche de modélisation continue, qui est utile pour la phase initiale de planification et de modélisation à grande échelle dans les grandes études régionales. L'objectif de cette approche est d'obtenir la tendance générale et le schéma de la distribution et du choix de voyage des utilisateurs. L'objectif est aussi l'étude des modifications de ces deux facteurs en réponse aux changements de politique dans le système de transport à l'échelle macroscopique, plutôt qu'une description détaillée du réseau. L'hypothèse fondamentale est que la différence dans les caractéristiques de modélisation, comme le coût du voyage et la demande entre zones adjacentes au sein d'un réseau, est relativement faible par rapport à la variation de l'ensemble du réseau. Par conséquent, les caractéristiques d'un réseau, telles que l'intensité de flux, la demande, et le coût du voyage, peuvent être représentés par des fonctions mathématiques régulières.

Dans la première partie, nous nous sommes concentrés sur les réseaux *ad hoc* sans fil, où nous avons considéré une fonction de coût générique, qui peut prendre en compte différents paramètres tels que la congestion du réseau, la quantité de nœuds de relais nécessaires

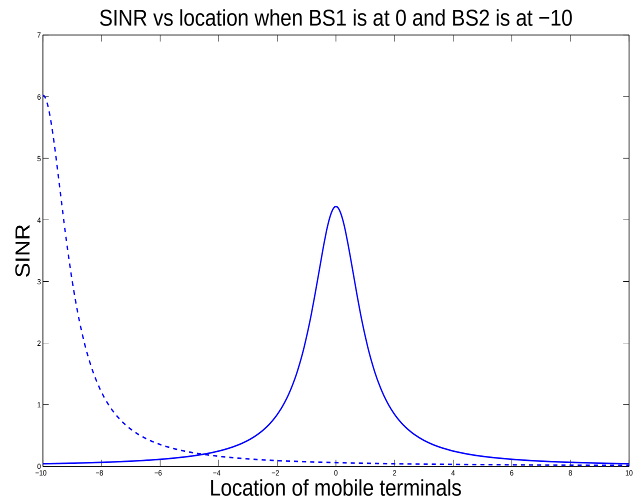


Figure 10: SINR comme fonction de l'ubication de terminal mobiles quand BS_1 est dans la position 0 (ligne plaine) et BS_2 dans -10 (ligne en pointillés).

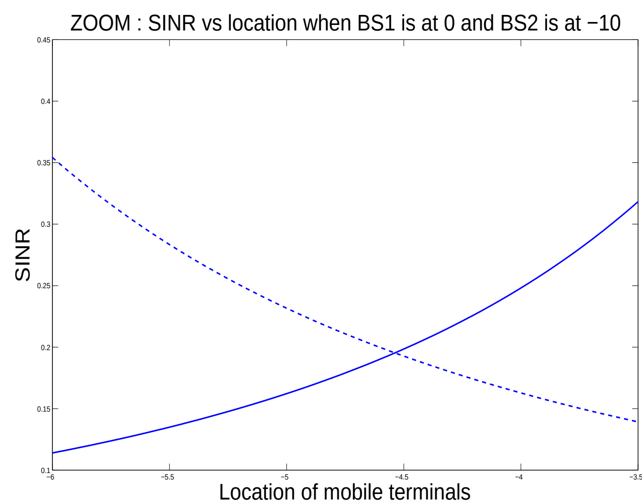


Figure 11: Zoom du SINR comme fonction de l'ubication de terminal mobiles quand BS_1 est dans la position 0 (ligne plaine) et BS_2 dans -10 (ligne en pointillés). Le meilleur équilibre est $\mathrm{eq}_1 = -4.68$ avec un valeur de SINR de 0.0025.

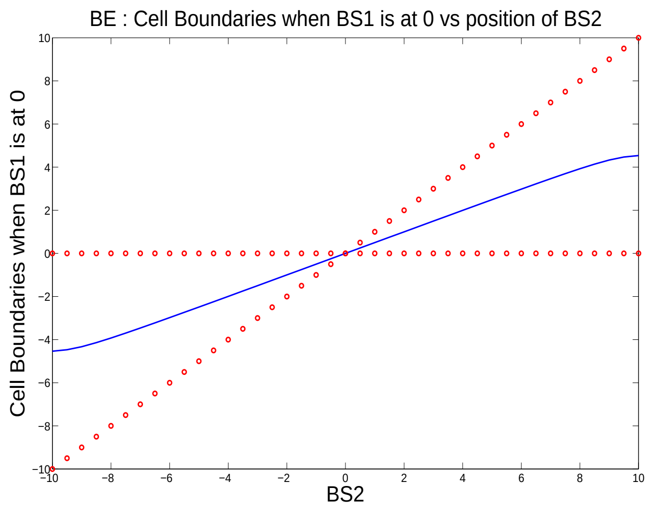


Figure 12: Meilleurs Équilibres: Seuils qui détérminent les frontières des cellules (dans l'axe vertical) comme fonction de l'ubication de BS_2 avec BS_1 dans la position 0.

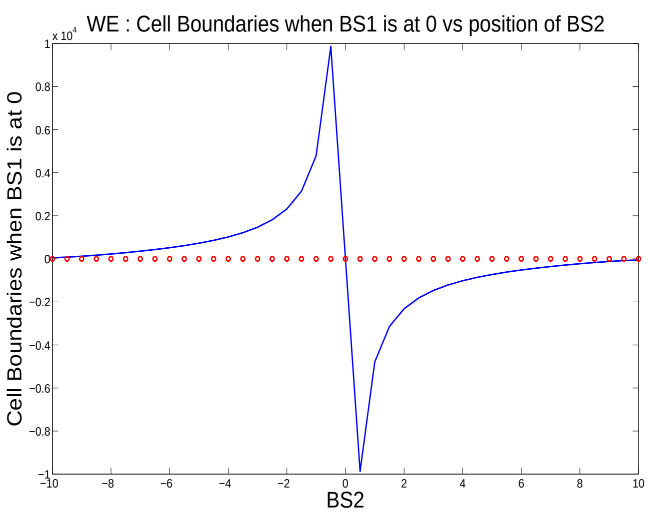


Figure 13: Pires Équilibres: Seuils qui détérminent les frontières des cellules (dans l'axe vertical) qui nous donnent le pire équilibre (en considerant le SINR comme utilité) comme fonction de l'ubication de BS_2 avec BS_1 dans la position 0.

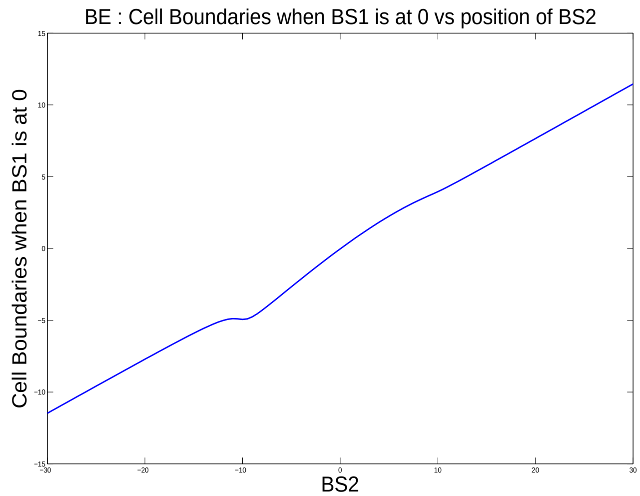


Figure 14: Le cas non-homogène: Seuils qui détérminent les frontières des cellules (dans l'axe vertical) des meilleurs équilibres (en considerant le SINR comme utilité) comme fonction de l'ubication de BS_2 avec BS_1 dans la position 0 quand nous considerons la distribution donné par $\lambda(x) = (L - x)/2L^2$.

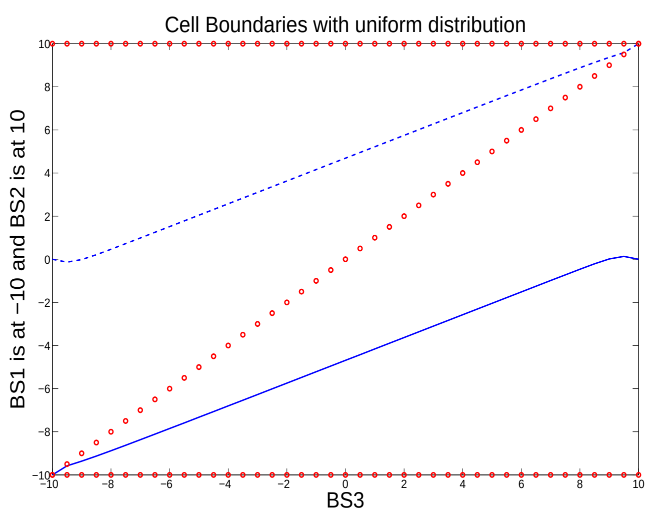


Figure 15: Plusiers BSs: Seuils qui détérminent les frontières des cellules (dans l'axe vertical) comme fonction de l'ubication de BS_3 pour $BS_1 = -10$ et $BS_2 = 10$.

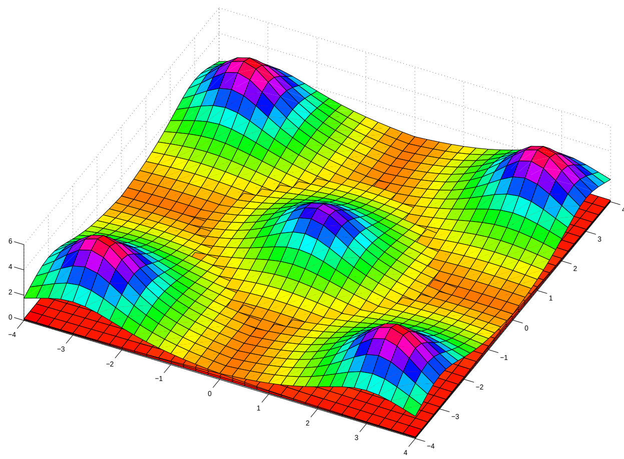


Figure 16: Le cas en 2D: Les frontières des cellules avec une distribution uniforme d'utilisateurs.

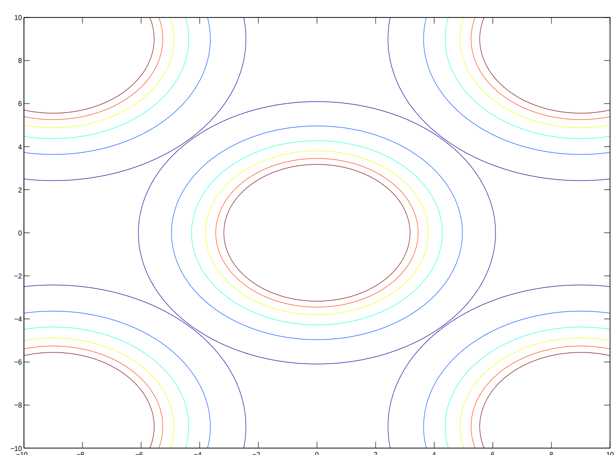


Figure 17: Le cas en 2D: Les contours des cellules avec une distribution uniforme d'utilisateurs.

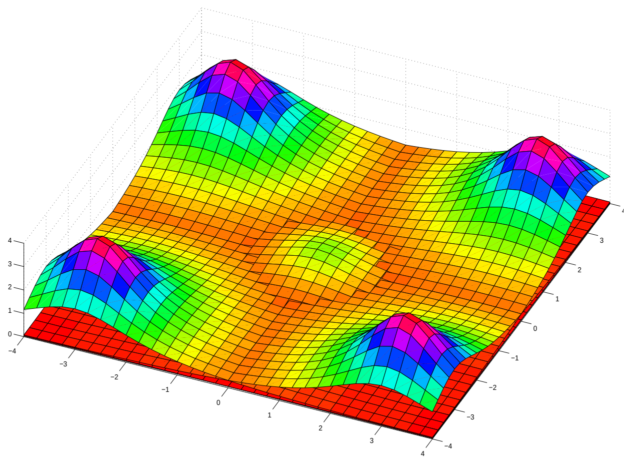


Figure 18: Le cas non-homogène en 2D : Les frontières des cellules avec une distribution non-uniforme d'utilisateurs.

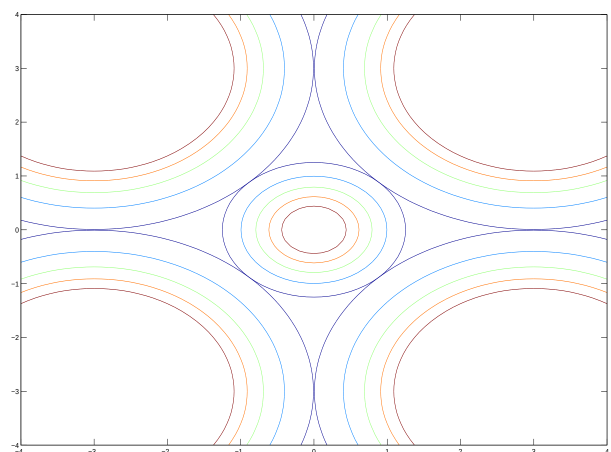


Figure 19: Le cas non-homogène en 2D : Les contours des cellules avec une distribution non-uniforme d'utilisateurs.

pour maintenir un certain débit, ou des paramètres liés à la consommation d'énergie du réseau. Avec cette fonction de coût, nous sommes en mesure de formuler et de résoudre le problème de routage de l'utilisateur et le problème d'optimisation du système pour les antennes directionnelles et les antennes omnidirectionnelles. Nous avons également trouvé des caractérisations du chemin de coût minimum grâce au théorème de Green. Lorsque nous permétons la mobilité des nœuds d'origine, et les nœuds de destination, nous sommes en mesure de donner un cadre pour modéliser et résoudre la quantité optimale de nœuds de relais nécessaires pour transmettre une certaine quantité de données.

Dans la deuxième partie, nous nous sommes concentrés sur les réseaux cellulaires, où nous avons étudié la capacité des canaux de diffusion MIMO (et des canaux de diffusion point-à-point MIMO) et le problème d'association de mobiles dans les réseaux cellulaires. En vertu de ce cadre, nous avons pu assurer la qualité de service, tout en minimisant la puissance totale du réseau. Nous avons résolu dans ce contexte le problème d'optimisation de l'utilisateur et le problème d'optimisation du système.

Perspectives

Il existe un certain nombre de problèmes ouverts qui peuvent être vus dans ce manuscrit. Toutefois, l'un des problèmes fondamentaux qui reste ouvert est de rompre la frontière entre la modélisation discrète et la modélisation continue. Pour étudier ce problème, nous pensons construire un cadre théorique cohérent pour aborder ces deux problèmes ensembles et analyser la convergence du problème discret vers le problème continu. Dans la même perspective, il serait d'un intérêt particulier d'étudier le problème d'optimisation du point de vue de l'utilisateur et analyser sa convergence.

Pour moi le passé ne passe pas. Il ne passe pas, n'est pas passé et ne passera jamais.
Quand on me dit "je suis quelqu'un de mon temps", on peut le dire, mais ce n'est pas vrai.
Et quelqu'un de son temps est quelqu'un de tous les temps aussi, du temps qui s'est écoulé.
Parce que la culture, la langue, l'histoire, tout ce qu'elle a trouvé à sa naissance
n'est pas son œuvre, c'est œuvre de toutes les générations qui l'ont précédé.
Alors la vérité c'est qu'il y a un pourcentage d'héritage qui se poursuit,
c'est-à-dire que nous héritons en permanence tout ce qui vient du passé :
le bien, le mal, le merveilleux, le pire et l'horrible.
Nous sommes les héritiers de tout ce qui a précédé et nous contribuons,
avec notre part, à l'héritage de ceux qui viendront après nous.
C'est une sorte de vague qui ne s'arrête jamais, qui va et vient en permanence
et qui accumule chaque fois plus de passé. Le passé augmente.

José Saramago.

Introduction

The explosive growth of wireless systems coupled with the proliferation of wireless devices such as 3G phones, WiFi laptops, and wireless sensor devices, have raised many technical challenges in the planning and analysis of wireless networks. Roughly speaking, wireless networks can be classified into two categories: wireless access networks and wireless ad hoc networks. On the one hand, wireless access networks usually provide connectivity to the wired infrastructure network through the wireless medium. Examples of wireless access networks include cellular systems such as 2G GSM, 3G UMTS, etc., and WiFi systems such as 802.11 WLANs. On the other hand, wireless ad hoc networks are decentralized wireless networks that may be set up and dismantled dynamically.

In this thesis, we are interested in massively dense wireless networks in cellular systems and massively dense wireless ad hoc networks. The term "massively dense" will be formally defined afterward. This thesis starts with the idea of studying wireless networks by analogy to what was called in a broad sense "physical tools". When we started this work, we discovered that there was a large amount of work being done and still under construction by another community: the road traffic community. We will briefly explain what this community is interested in and why this is interesting for our work on the modeling of wireless networks. In the literature of the road traffic community the modeling of traffic equilibrium problems for a transportation system is classified in the discrete modeling approach and the continuum modeling approach:

- In the discrete modeling approach, each road link in the network is modeled separately and the demand is assumed to be concentrated at hypothetical points, called zone centroids. This modeling approach is commonly adopted for the detailed planning and analysis of transportation systems.

- The continuum modeling approach is used for the initial phase of planning and modeling in broad-scale regional studies. In this setting, the focus is on the general trend and pattern of the distribution, the travel choice of users, and on changes in these two factors in response to policy changes in the transportation system. The fundamental assumption is that the difference in modeling characteristics, such as the travel cost and the demand pattern between adjacent areas in the network, is relatively small compared to the variation over the entire network. Hence, the characteristics in a network, such as the flow intensity, demand, and travel cost, can be represented by smooth mathematical functions [40].

The continuum modeling approach has many advantages over the discrete approach in macroscopic studies on dense transportation systems. First, it reduces the problem size for dense transportation networks. The problem size in the continuum model depends on the method that is adopted to approximate the modeling region, but not on the actual network itself. Because of that, an effective approximation method, such as the finite element method (FEM), can extensively reduce the size of the problem. This reduction in problem size saves computational time and memory. Second, less data is required to model the set-up in a continuum model. As continuum modeling can be characterized by a small number of spatial variables, it can be set-up with a much smaller amount of data than the discrete modeling approach, which requires data for all of the included links. This makes the continuum model convenient for macroscopic studies in the initial phase of design since the collection of data in this phase is time consuming and labor intensive, and the resources to undertake it are generally not available, which means there is usually insufficient data on the system to set up a detailed model. Finally, the continuum modeling approach gives a better understanding of the global characteristics of a network. To our surprise the advances of this community are not well known in our community. In this context we first analyze the routing problem in wireless ad hoc sensor networks which has to deal with the transportation of packets through the network. The topic of transportation has been addressed earlier in the context of optimal allocation of resources through linear programming by Hitchcock [3] and Kantorovich [4] (who later shared the Nobel Prize with Koopmans) as well as by Koopmans [5] and Dantzig [6]. In such models, however, there was no congestion associated with transportation. In 1952, Wardrop [2] had set two principles of transportation network utilization, which have become to be termed, respectively user-optimization and system-optimization:

- The first principle expresses that vehicles select their routes of travel from origins to destinations independently. Then, in an equilibrium situation, the journey times of all routes actually used between an origin/destination pair are equal. This journey time is less than the journey time that would be experienced by a single vehicle on any unused route. The user-optimized solution is also referred to as traffic network equilibrium or as traffic assignment.

- The second principle reflects the situation in which there is a central controller which routes the traffic flows in an optimal manner from origins to destinations to minimize the total cost of the network.

In 1956, Beckmann, McGuire, and Winsten [7] were the first to provide a rigorous mathematical formulation of the conditions set forth by Wardrop's first principle that allowed for the ultimate solution of the traffic network equilibrium problem in the context of certain link cost functions which were increasing functions of the flows on the links. In particular, they demonstrated that the optimality conditions in the form of Karush-Kuhn-Tucker [8, 9] conditions of an appropriately constructed mathematical programming/optimization problem coinciding with the statement that the travel costs on utilized routes/paths connecting each origin/destination pair of nodes in a transportation network have equal and minimal travel costs. Hence, no traveler, acting unilaterally will have any incentive to alter its path (assuming rational behavior) since his travel cost (travel time) is minimal. Thus, a problem in which there are numerous decision-makers acting independently and as later also noted by Dafermos and Sparrow [10] competing in the sense of Nash [1], could be reformulated (under appropriate assumptions that will be defined later) as a convex optimization problem with a single objective function subject to linear constraints and non-negativity assumptions on the flows in the network.

Beckman [11] noted the relevance of network equilibrium concepts to communication networks. In another interesting work, Bertsekas and Gallager [12] realized the similarities between communication and transportation networks. The work on the Braess paradox [13] subsequently provided one of the main links between transportation science and computer science. In 1990, Cohen and Kelly [14] described a paradox analogous to that of the Braess in the case of a queueing network. The Braess paradox still continues to be investigated in the road-traffic context [15, 16], as well as in the networking community [17, 18, 19].

In the networking community there have been some problems in the analysis of what has received the name of massively dense networks. When the network has an increasing number of nodes, the modeling and analysis of the network is much more difficult and sometimes intractable. When we speak about dense networks, we assume a strong separation in spatial scales between the macroscopic level, corresponding to typical distances between the source and destination nodes, and the microscopical level, corresponding to distances between the neighboring nodes. When the system is sufficiently large, the macroscopic model will give a better description of the network and we can derive its properties from microscopic considerations. We will sacrifice a detailed description of the optimized solution but the macroscopic model will preserve enough information in order to give a good description of the network and the derivation of insightful results under different settings.

The physics-inspired methodologies used for the study of dense ad hoc networks starts with the pioneering works of Jacquet [20], and Kalantari and Shayman [21, 22]. In this area, a number of research groups have worked on dense ad hoc networks using tools from Geometrical Optics [20] as well as Electrostatics (see e.g. [23, 24, 25], and the survey [26] and references therein). We shall describe these in the next sections, with particular attention in Chapter 1, Part I. The physical paradigms allow the authors to minimize various metrics related to the routing. Hyytia and Virtamo propose in [27] an approach based on load balancing arguing that if shortest path (or cost minimization) arguments were used, then some parts of the network would carry more traffic than others and may use more energy than others. This would result in a shorter lifetime of the network since some parts would be out of energy earlier than others and earlier than any part in a load balanced network.

The development of the original theory of routing in massively dense networks among the community of ad hoc networks has emerged in a complete independent way of the existing theory of routing in dense networks which had been developed within the community of road traffic engineers. The approach introduced in 1952 by Wardrop [2] and by Beckmann [28] is still an active research area among this community, see [29, 30, 31, 32, 33] and references therein. We combine in this thesis various approaches from this area as well as from optimal control theory and optimal transportation theory in order to formulate models for routing in massively dense networks.

The main contributions of the thesis will be the optimal deployment of relay nodes in the case of massively dense wireless ad hoc networks and the optimal deployment of base stations and mobile association in the case of wireless access networks.

Thesis Organization and Contribution

Part I - Optimal Planning for Massively Dense Wireless Ad Hoc Networks:
Routing optimization

1. Basic Concepts on Massively Dense Ad Hoc Networks

   This chapter gives an introduction to the main problems related to the optimal routing and analysis of massively dense ad hoc network in the user- and system-optimization context. We consider a generic cost function and we also describe some important particular cases for this function. The objective is to understand the context of the optimal routing and the scalability problem within this type of networks.

2. Electrostatics Approach

   This chapter first gives a brief summary of the related work and describes the main approaches that have been made by doing a parallel to Optics and Electrostatics in the optimal placement of relay nodes for wireless static ad hoc networks. We start by an example to describe the problem in the one-dimensional case and then the extension of the results for the two-dimensional case.

3. Directional Antennas

   This chapter focus on directional antennas, and gives an extension of the concept of Wardrop equilibrium for the continuum case. We are able to show a simple path characterization (under appropiate assumptions) by the use of Green's theorem.

4. Omni-directional Antennas

   This chapter extends the previous results for the user- and system-optimization problem in the context where the mobiles can choose any direction at any given time.

5. Numerical Analysis

   In many cases, we have that the optimal solution to both the user- and the system-optimization problem are given by a partial differential equation. We propose the finite elements method to solve this equation since this method allow us to give bounds in the variation of the solution with respect to the variation of the data in the considered problem.

6. Magnetworks: Mobility of the nodes

In this chapter we give the results related to the mobility of the nodes. Even when we allow mobility of the origin and destination nodes, we are able to find the optimal quantity of relay nodes needed to support a certain throughput. We also show a relation between the metrics of Electrostatics and Optics works.

Part II. - Optimal Planning of Cellular Networks

7. Capacity of Networks with MIMO capabilities

   This chapter gives an introduction to random matrix theory and its application to Network MIMO (multiple-input and multiple-output). We have shown that even when the channel offers an infinite number of degrees of freedom, the capacity is mainly limited by the ratio between the size of the antenna array at the base station and the mobile terminals and the wavelength of the signal.

8. Mobile association and optimal placement of base stations

   This chapter gives an introduction to optimal transport: Monges problem, duality results from Kantorovich and some of the consequences. It presents the solution to the quality of service (QoS) problem in the downlink and uplink case results where different policies are analyzed.

- Summary and Conclusion

  This chapter presents a summary of the main results of the thesis and the conclusions that can be extracted from them. This chapter gives an insight new perspectives and future works that can be devised from the results of this thesis.

Published Works

International Journal Papers

- "Continuum equilibria and global optimization for routing in dense static ad hoc networks," A. Silva, E. Altman, P. Bernhard, M. Debbah, Computer Networks, In Press, Corrected Proof. ISSN: 1389-1286, DOI: 10.1016/j.comnet.2009.10.019

- "Optimal mobile association in dense cellular networks," A. Silva, H. Tembine, Ch. Jimenez, E. Altman, M. Debbah, To be submitted.

Papers in Proceedings of Refereed Conferences

- "Spatial Games combining base station placement and mobile association: the downlink case," A. Silva, H. Tembine, E. Altman, M. Debbah, Submitted to CDC'2010: 49th IEEE Conference on Decision and Control, Atlanta, Georgia, December 15-17, 2010.

- "Magnetworks: how mobility impacts the design of mobile ad hoc networks," A. Silva, E. Altman, M. Debbah, G. Alfano, Proceedings of IEEE INFOCOM 2010, San Diego, CA, USA, March 15-19, 2010.

- "Congestion in Randomly Deployed Wireless Ad Hoc Sensor Networks," A. Silva, P. Reyes, M. Debbah, Proceedings of International Conference on Ultra Modern Telecommunications, St-Petersburg, Russia, October 12-14, 2009.

- "Stochastic Games with One-Step Delay Sharing Information Pattern with Application to Power Control," E. Altman, V. Kamble, A. Silva, Proceedings of Gamenets 2009, Istanbul, Turkey, May 13-15, 2009.

- "Numerical Solutions of Continuum Equilibria for Routing in Dense Ad-Hoc Networks," A. Silva, P. Bernhard, E. Altman, Proceedings of Valuetools 2008, Workshop Inter-Perf, Athens, Greece, October 20-24, 2008.

- "The Space Frontier: Physical Limits of Multiple Antenna Information Transfer," R. Couillet, S. Wagner, M. Debbah, A. Silva, Proceedings of Valuetools 2008, Workshop Inter-Perf, Athens, Greece, October 20-24, 2008.
  Best student paper award.

- "The Mathematics of Routing in Massively Dense Ad-Hoc Networks," P. Bernhard, E. Altman, A. Silva, Proceedings of AdHoc-NOW Conference, Sophia-Antipolis, France, September 10-13, 2008.

- "Continuum Equilibria in Spatial Routing Games," A. Silva, E. Altman, M. Debbah, P. Bernhard, Invited Paper, *9ème congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision*, Clermont-Ferrand, France, February 25-28, 2008.

- "Continuum Equilibria for Routing in Dense Ad-Hoc Networks," E. Altman, A. Silva, P. Bernhard, M. Debbah, Proc. 45th Annual Allerton Conference on Communication, Control, and Computing, Illinois, USA, September 26-28, 2007.

Part I

Optimal Planning for Massively Dense Wireless Ad Hoc Networks

Routing optimization for massively dense wireless ad hoc networks

In this part of the thesis, we are interested in the routing optimization problem in massively dense ad hoc networks. We also consider the problem of the optimal deployment of nodes in the network by taking into account the optimal routing.

The routing optimization problem can take into account different cost functions and each cost function may produce a different solution. Moreover, the cost function could take into account several network metrics. Because of this, we start by considering a general cost function. However, when the particular structure of the network allow us to provide some insightful solution to the problem, we consider some particular cost function. Within this context, we restrict our attention to the routing problem in massively dense ad hoc networks. These networks can be considered as fluid approximations of ad hoc networks which have a large number of nodes. We combine several approaches, from the area of road traffic engineering as well as from optimal control theory, to analyze and solve this problem in different situations.

We start by providing several cost functions included in the scope of our work. We give a brief description of related works and their contribution to the routing problem in massively dense ad hoc networks. To introduce the subject, we present the problem in the one-dimensional setting, where we are able to find the deployment of nodes needed to maintain the optimal throughput at each portion of the network. Then, we focus on the two-dimensional setting in the case of directional antennas and omni-directional antennas in static ad hoc networks. We analyze the routing problem in directional antennas when only horizontal and vertical transmissions are allowed and in omni-directional antennas when any direction can be chosen at any location. The fluid approximation methodology allow us to solve different optimization problems: the system-optimization problem in which the objective is to minimize the total cost of the network, and the user-optimization problem, where each user seeks to minimize its own cost function. In the latter case, we analyze the equilibrium situation, that we denote Wardrop equilibrium by analogy with the work developed in [2]. A simple characterization of minimum cost paths for the user-optimization problem can be found, by using a classical control based on Green's theorem. As in many situations the solution to the optimization problem is characterized by a partial differential equation, we propose the numerical analysis of these equations by finite elements methods. We briefly explain this method and provide bounds for this numerical approximation. Afterwards, we extend our results to mobile ad hoc networks where we consider different mobility scenarios.

Chapter 1

Basic Concepts on Massively Dense Wireless Ad Hoc Networks

In this chapter we give a brief introduction to massively dense wireless ad hoc networks. We describe the physics-inspired methodologies that have been used to analyze the optimal planning and analysis of these networks taking into account the routing problem in the user-optimization and system-optimization context. We present Road Traffic theory in this context and the tools that have been used to analyze this type of networks. We give a general framework for the cost function but we also present some particular cost functions related to the congestion of the network, the capacity scaling of the network, and costs related to energy consumption.

1.1 Introduction

We are interested in the modeling and analysis of massively dense wireless ad hoc networks. First question to ask is: What are wireless ad hoc networks? Wireless ad hoc networks are basically decentralized ad hoc networks. This type of wireless networks is called "ad hoc" because it doesn't depend on a preexisting infrastructure. In wired networks, routers control and operate the network. Access points do the same in managed wireless networks. Instead, in wireless ad hoc networks each node participates in the routing process by forwarding data for other nodes, and so the determination of which nodes forward data is made dynamically based on the network connectivity. Wireless ad hoc networks have many advantages with respect to wired networks such as lower installation and maintenance costs, ease of replacement and upgrading, reduced connector failure, greater physical mobility, etc. Minimal configuration and quick deployment make ad hoc networks suitable, for example, for emergency situations like natural disasters or military conflicts. For a nice introduction to the advantages of wireless ad hoc networks and its industrial applications see *e.g.* [41].

Research on wireless ad hoc networks involves many issues such as the design of protocols at various network layers (MAC, transport, etc.), the investigation of physical limits of transfer rates, the optimal design of end-to-end routing, efficient energy management, connectivity and coverage issues, performance analysis of delays, loss rates, etc. The study of these issues has required the use of engineering methodologies as well as information theoretical ones, control theoretical tools, queueing theory, and others. One of the most challenging problems in the performance analysis and in the control of ad hoc networks has been the routing problem. There are two approaches to this problem: the discrete and the continuous modeling approach. The discrete modeling approach for this problem consist on considering a graph, *i.e.*, a set of vertices (or nodes) and a set of edges (or the set of possible transmissions between nodes, when by possible transmission we mean that the receiving node is on the transmission range of the transmitting node), and a real-valued weight function which takes into account the transmission cost between the transmitting and receiving nodes for every single transmission. One of the main problems analyzed within this context is the shortest path problem.

The shortest path problem or minimum cost path is the problem of finding a path between two vertices (or nodes) such that the sum of the weights (or transmission cost between nodes) of its constituent edges is minimized. Then the objective is to find a path from the origin to the destination node such that the total cost of the path (considered as the sum of the transmission costs of every transmission between the nodes that belong to the path) is minimal among all paths connecting the origin to the destination. The most important algorithms to deal with this problem are:

- Dijkstra's algorithm which solves the single-pair, single-source, and single-destination shortest path problem, which by using Big-O notation has running time of $\mathrm{O}(\text{number of vertices}^2)$.
- Bellman-Ford algorithm which solves the single source problem when the edge weights may be negative, which has running time of $\mathrm{O}(\text{number of vertices} \times \text{number of edges})$.
- Floyd-Warshall algorithm (sometimes known as the WFI algorithm or Roy-Floyd algorithm) which solves all pairs of shortest paths and has running time of $\mathrm{O}(\text{number of vertices}^3)$.

For an explanation on these algorithms see *e.g.* [34].

The problem with these algorithms is that they need an exponential running time with respect to the number of entries and that the quantity of data required by the discrete modeling may not be available, *e.g.* for the initial phase of planning and modeling in broad-scale regional studies there are mainly estimations but not precise data; or our focus in on the general trend and pattern of the distribution and travel choice of packets, and on changes in response to policy changes in the network. The other problem with just solving and implementing the shortest path problem was given by Hyytiä and Virtamo in [27]. As previously mentioned, they proposed an approach based on load balancing arguing that if shortest path (or cost minimization) arguments were used, then some parts of the network would carry more traffic than others and would use more energy than others and this would result in a shorter lifetime of the network since some parts would be out of energy earlier than others and earlier than any part in a load balanced network. On the other hand, the continuum modeling approach used for obtaining the general trend has various advantages over the discrete approach, since it reduces the problem size for dense transportation networks given that the problem size in the continuum model depends only on the method that is adopted to approximate the modeling region but not on the actual network itself. The fundamental assumption on the continuum modeling approach is that the difference in modeling characteristics, such as the travel cost and the demand pattern between adjacent areas within the network, is relatively small compared with the variation over the entire network. Hence, the characteristics in a network, such as the flow intensity, demand, and travel cost, can be represented by smooth mathematical functions.

We will work on the continuum modeling of massively dense wireless ad hoc networks. The term "massively dense" is used to indicate not only that the number of nodes in the network is large, but also that the network itself is highly connected. By the term "dense" we further understand that for every point in the plane there is a node close to it with high probability; by "close" we mean that its distance is much smaller than the transmission range. Another possibility to describe this type of networks is to assume a strong separation in spatial scales between the macroscopic level, corresponding to typical distances between the source and destination nodes, and the microscopic level, corresponding to distances between the neighboring nodes.

The main problem on dealing with massively dense ad hoc networks is that when applying existing tools for optimal routing, the complexity makes the solution intractable as the number of nodes becomes very large. Nevertheless, it has been observed that as an ad hoc network becomes "more dense" (in a sense that will be defined precisely later), the optimal routes seem to converge to some limit curves. This is illustrated in Fig. 1.1. We call this regime, the limiting "macroscopic" regime. We shall show that the solution to the macroscopic behavior (*i.e.*, the limit of the optimal routes as the system becomes more and more dense) is sometimes much easier to solve than the original "microscopic model".

The empirical discovery of the macroscopic limits motivated a large number of researchers to investigate continuum-type limits of the routing problem. A very basic problem in doing so has been to identify the most appropriate scientific context for modeling and solving this continuum limit routing problem. Our major contribution in this part of the thesis is to identify the main paradigms (from optimal control as well as from road traffic engineering) for the modeling and the solution of this problem.

1.1.1 Physics-inspired paradigms

The physics-inspired paradigms used for the study of massively dense ad hoc networks go way beyond those related to statistical-mechanics in which macroscopic properties are derived from microscopic structure. Starting from the pioneering work of Jacquet [20], and Kalantari and Shayman [21, 22], a number of research groups have worked on massively dense ad hoc networks using tools from geometrical Optics [20], as well as Electrostatics [21, 22]. Popa *et al.* studied in [42] optical paths and actually showed that the optimal solution to a min-max problem of load balancing can be achieved by using an appropriately chosen optical profile. The forwarding load corresponds to the scalar sum of traffic flows of different classes. This means that the optimal solution (with respect to this objective) can be achieved by single path routes, a result obtained in [42, 43]. Similar problems have been also studied in [44], as well as in works doing load balancing by analogies to Electrostatics (see

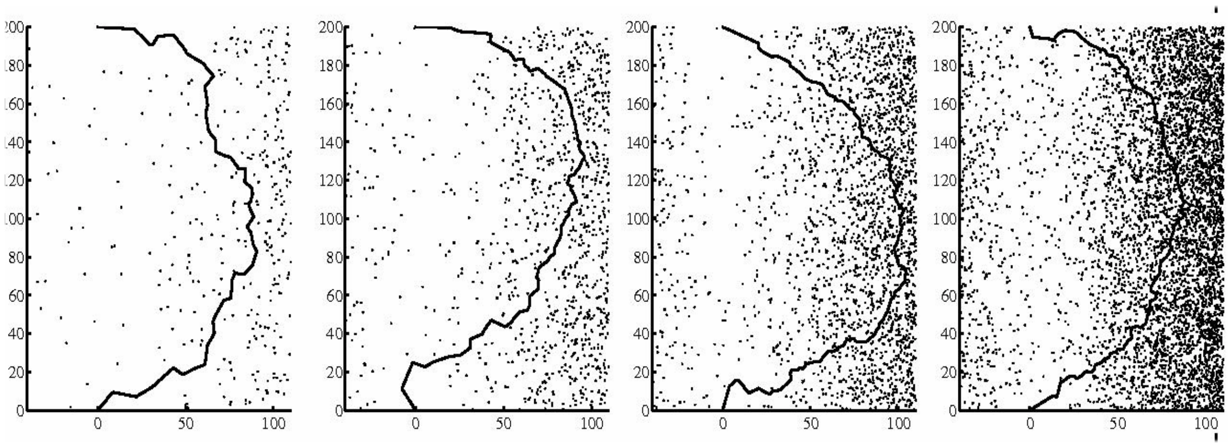


Figure 1.1: Minimum cost routes connecting a source node placed at the origin $(0\text{ m}, 0\text{ m})$ and a destination node placed at the location $(0\text{ m}, 200\text{ m})$, through an area where relay nodes are placed according to a spatial Poisson process of density $\lambda(x_1, x_2) = a \times [10^{-4}x_1^2 + 0.05]$ nodes per m^2, for four increasing values of a $(a = 1/30, 1/10, 1/5, 1/2)$ in increasingly large networks. This figure was gently given by Prof. Toumpis.

e.g. [21, 22, 23, 24, 25], and the survey [26] and references therein). The physical paradigms allow the authors to minimize various metrics related to the routing problem. Hyytia and Virtamo proposed in [27] an approach based on load balancing arguing that if shortest path (or cost minimization) arguments were used, then some parts of the network would carry more traffic than others and may use more energy than others. This would result in a shorter lifetime of the network since some parts would be out of energy earlier than others.

1.1.2 Note on road traffic theory

The development of the original theory of routing in massively dense networks among the community of ad hoc networks has emerged in a completely independent way of the existing theory of routing in massively dense networks, which has been developed within the community of road traffic engineers. Indeed, this approach has already been introduced in 1952 by Wardrop [2] and Beckmann [28] and is still an active research area among that community (see *e.g.* [29, 30, 32, 33], the survey [31], and references therein).

1.2 Cost functions for the routing problem

In optimizing a routing protocol in ad hoc networks, or in optimizing the placement of nodes, one of the starting points is the determination of the cost function that captures the cost of transporting a packet through the network. To determine it, we need a specification of the network which includes the following:

- A network topology for the network.
- A forwarding rule that nodes will use to select the next hop of a packet.
- The cost incurred for transmitting a packet to an intermediate node.

Below we present several ways of choosing cost functions.

We define the flow of information $\mathbf{T}(\mathbf{x})$ (see Fig. 2.4) to be a vector whose components are the horizontal and vertical flows at location $\mathbf{x}$. Throughout we assume that each point carries a single flow (although the methodology can be extended to the multi-class flow case). The restriction to a single flow is justified when there is either a single destination, or when there is a set of destination points and the routing protocol has the freedom to decide to which of the set the packets will be routed. Under this type of conditions, one may assume a single flow at each point without loss of optimality (see *e.g.* [43]).

In this section we present several cost functions. We analyze the problem in its more general form by considering a general cost function, and then we will further investigate some particular cost functions that are of our interest.

1.2.1 Costs related to congestion

Congestion independent routing

A metric often used in the Internet for determining routing costs is the number of hops from origins to destinations, which routing protocols try to minimize. The number of hops is proportional to the expected delay along the path in the context of ad hoc networks, in case the queueing delay is negligible with respect to the transmission delay over each hop. This criterion is insensitive to interference or congestion. We assume that it depends only on the transmission range. We describe various cost criteria that can be formulated with this approach.

- If the range is constant then the cost density $c(\mathbf{x})$ is constant so that the cost of a path is its length in meters. The routing then follows a shortest path selection.

- Let us assume that the range $R(\lambda; \mathbf{x})$ is small, and it depends on local radio conditions at position $\mathbf{x}$ (for example, if it is influenced by weather conditions) but not on interference. The latter is justified when dedicated orthogonal channels (*e.g.* in time or frequency) can be allocated to traffic flows that would otherwise interfere with each other. Then determining the optimal routing becomes a path cost minimization problem. We further assume, as in [45], that the range is scaled to go to 0 as the total density λ of nodes grows to infinity. More precisely, let us consider a scaling of the range such that the following limit exists:

$$r(\mathbf{x}) := \lim_{\lambda \to \infty} \frac{R(\lambda; \mathbf{x})}{\lambda}$$

  Then in the dense limit, the fraction of nodes that participate in forwarding packets along a path is $1/r(\mathbf{x})$ at position $\mathbf{x}$, and the path cost is the integral of this density along the path.

- The influence of varying radio conditions on the range can be eliminated using power control that can equalize the hop distance.

Congestion dependent routing

Another more general transmission cost function may depend on a measure of the congestion of the network. For instance a transmission cost c may depend on the traffic flow $\mathbf{T}$ that is passing through a particular location $\mathbf{x}$. In this case we will assume that it will depend on the magnitude of the traffic flow $\|\mathbf{T}(\mathbf{x})\|$ at that location but not in its direction. And even more, it may depend on the location at which the transmission takes place. In which case, $c = c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)$.

1.2.2 Costs derived from capacity scaling

One particular class of cost functions is given by the quantity or density of nodes needed to maintain the traffic requirements of the network. In that framework, it makes sense to investigate how much traffic can be carried by a certain quantity or density of nodes. That quantity receives the name of transport capacity. Many models have been proposed in the literature that show how the transport capacity scales with the number of nodes n or with the density of nodes λ within a certain region. Then the typical cost (see *e.g.* [24]) considered at a neighborhood of a location[1] $\mathbf{x}$ is the density of nodes required there to carry a given flow of information $\mathbf{T}(\mathbf{x})$. We work within a general framework and then investigate some particular cases with different protocols. Assume that we use a protocol that provides a transport capacity of the order of $f(\lambda)$ at some region in which λ denotes the density of nodes within that region (we will provide examples for function f ahead). This means that in order to support a flow of information of norm $\|\mathbf{T}(\mathbf{x})\|$ passing through a neighborhood of the location $\mathbf{x}$, we need to place deterministically the nodes according to

[1] We denote the vectors by bold fonts.

the formula $f^{-1}(\|\mathbf{T}(\mathbf{x})\|)$. Then if we assume that a flow of information $\mathbf{T}(\mathbf{x})$ is assigned to location $\mathbf{x}$, the cost will be taken as

$$c(\mathbf{x}, \mathbf{T}(\mathbf{x})) = f^{-1}(\|\mathbf{T}(\mathbf{x})\|) \tag{1.1}$$

where $\|\cdot\|$ represents the norm of a vector. As our protocol provides a transport capacity of the order of $f(\lambda)$ and our cost is the density. Then $f \circ f^{-1}(\mathbf{T}(\mathbf{x})) = \mathbf{T}(\mathbf{x})$ if the function f is invertible. Most of the works in this area has considered the ℓ^2-norm, *i.e.*, for $\mathbf{x} = (x_1, x_2)$, we define $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$.

Examples for function f:

- Using a network theoretic approach based on multi-hop communication, Gupta and Kumar proved in [45] that the throughput of the system that can be transported by the network when the nodes are optimally located is[2] $\Omega(\sqrt{\lambda})$, and when the nodes are randomly located this throughput becomes $\Omega(\sqrt{\lambda}/\sqrt{\log \lambda})$. Using percolation theory, the authors of [46] have shown that in the randomly located set the same $\Omega(\sqrt{\lambda})$ can be achieved.
- Baccelli, Blaszczyszyn and Mühlethaler introduce in [47] an access scheme, denoted MSR protocol (Multi-hop Spatial Reuse Aloha), reaching the Gupta and Kumar bound which does not require prior knowledge of the node density.

For the model of Gupta and Kumar with either the optimal location or the random location approaches, as well as for the MSR protocol (Multi-hop Spatial Reuse Aloha) with a Poisson distribution of nodes, we obtain a quadratic cost of the form

$$c(\mathbf{T}(\mathbf{x})) = k\|\mathbf{T}(\mathbf{x})\|^2 = k(T_1(\mathbf{x})^2 + T_2(\mathbf{x})^2). \tag{1.2}$$

This follows from the fact that in the previous examples $f(x)$ behaves like $\sqrt{x}$, so the inverse of the function f must be quadratic. Then from (1.1) we conclude that the cost function derived from capacity scaling in the previously analyzed cases must be quadratic on $\|\mathbf{T}(x)\|$.

Toumpis and Tassiulas in [48] focus on a particular physical layer model characterized by the following assumption:

Assumption 1: A location $\mathbf{x}$, where the node density is $\eta(\mathbf{x})$, can support any traffic flow vector with a magnitude less or equal to a bound $\|\mathbf{T}(\mathbf{x})\|_{max}$ which is proportional to the square root of the density, *i.e.* $\|\mathbf{T}(\mathbf{x})\| \leq \|\mathbf{T}(\mathbf{x})\|_{max} = K\sqrt{d(\mathbf{x})}$.

The validity of Assumption 1 depends on the physical layer and the medium access control protocol used by the network. Although it is not generally true, it holds in many different settings of interest. For example, in [48] Toumpis and Tassiulas give an example of network where m^2 nodes are placed in a perfect square grid of $m \times m$ nodes and each node can listen to transmissions from its four nearest neighbors. They give a simple time division scheme so that the network of m^2 nodes can support a traffic on the order of m. In [49] it was shown

[2] We denote $f \in \Omega(g)$ if f is bounded below by g (up to a constant factor) asymptotically and we denote $f \in \Theta(g)$ if f is bounded both above and below by g (up to a constant factor) asymptotically.

that the traffic that can be supported in the above network, if nodes access the channel by use of slotted Aloha instead of time division, is $\mathbf{T}_{\text{local}} = K \times W \times m$, where nodes transmit data with a fixed global rate of W bps, K is a constant-smaller than 1/3 (that captures the efficiency of Aloha). Finally, in [45] it was shown that a network of n randomly placed nodes can support an aggregate traffic on the order of $\sqrt{n/\log n}$ under a more realistic interference model that accounts for interference coming from arbitrarily distant nodes. The logarithm in the denominator appears due to the methodology of [45], and it has been shown [46] that it can be removed by use of percolation theory.

1.2.3 Costs related to energy consumption

In the absence of capacity constraints, the cost can represent energy consumption. In a general multi-hop ad hoc network, the hop distance can be optimized so as to minimize the energy consumption. Even within a single cell of 802.11 IEEE wireless LAN one can improve the energy consumption by using multiple hops, as it has been shown not to be efficient in terms of energy consumption to use a single hop [50].

Alternatively, the cost can take into account the scaling of the nodes (as we have done in Section 1.2.2) that is obtained when there are energy constraints. As an example, assuming random deployment of nodes, where each node has data to send to another randomly selected node, the capacity (in bits per Joule) has the form $f(\lambda) = \Omega\left((\lambda/\log\lambda)^{(q-1)/2}\right)$ where q is the path-loss, see [51]. The cost is then obtained using (1.1).

Chapter 2

Electrostatics Approach

In the work of Toumpis *et al.* ([25, 48, 24, 23, 26, 52]), the authors addressed the problem of the optimal deployment of massively dense wireless ad hoc networks by analogy with Electrostatics. We shall recall below the representation of the flow conservation constraint, which is well known in Electrostatics. This derivation appears both in physics-inspired works, as well as in the road traffic literature [29].

We first consider the one dimensional case in order to explain the main concepts involved in our model and how these concepts can be extended to the two dimensional case in order to obtain the optimal deployment of the relay nodes in a wireless ad hoc network.

2.1 Fluid approximations: one-dimensional case

As a first approach we consider the line segment $[0, L]$ as the geographical reference of the network. We consider the continuous *node density function* $\eta(x)$, measured in nodes/m, such that the total number of nodes on a segment $[\ell_0, \ell_1]$, denoted by $N(\ell_0, \ell_1)$, is

$$N(\ell_0, \ell_1) = \int_{\ell_0}^{\ell_1} \eta(x)\, dx.$$

We consider as well the continuous *information density function* $\rho(x)$, measured in bps/m, generated by the nodes such that

- At location x where $\rho(x) > 0$ there is a fraction of data created by the sensor sources, such that the rate with which information is created in an infinitesimal area of size $d\varepsilon$, centered at position x, is equal to $\rho(x)\, d\varepsilon$.

- Similarly, at location x where $\rho(x) < 0$ there is a fraction of data received at the sensor destinations such that the rate with which information is received by an infinitesimal area of size $d\varepsilon$, centered at position x, is equal to $-\rho(x)\, d\varepsilon$.

We assume that the total rate at which sensor destinations have to receive data is the same as the total rate which the data is created at the sensor sources, *i.e.*,

$$\int_0^L \rho(x)\,dx = 0. \tag{2.1}$$

Notice that if we have an estimation of the packet loss through the network, we can put different weights to the evaluations of the function ρ in order to adequate the function to satisfy equation (2.1).

Consider the continuously differentiable *traffic flow function* $T(x)$, measured in bps/m, such that its direction (positive or negative) coincides with the direction of the flow of information at point x and $\|T(x)\|$ is the rate at which information propagates at position x, *i.e.*, $\|T(x)\|$ gives the total amount of traffic that is passing through the position x.

Next we present the flow conservation condition. In order to conserve the information transmitted over a line segment $[\ell_0, \ell_1]$, it is necessary that the rate with which information is created over the segment is equal to the rate with which information is leaving the segment, *i.e.*,

$$T(\ell_1) - T(\ell_0) = \int_{\ell_0}^{\ell_1} \rho(x)\,dx.$$

The integral on the right hand side is equal to the quantity of information generated (if it's positive) or demanded (if it's negative) by the fraction of nodes over the line segment $[\ell_0, \ell_1]$. The expression $T(\ell_1) - T(\ell_0)$, measured in bps/m, is equal at the rate with which information is leaving (if it's positive) or entering (if it's negative) the segment $[\ell_0, \ell_1]$. This holding for any line segment, it follows that necessarily,

$$\frac{dT(x)}{dx} = \rho(x) \quad \text{for all } x \in (0, L). \tag{2.2}$$

The problem considered is to find the number of nodes $N(0, L)$ in the line segment $[0, L]$, needed to support the information created by the sources and received at the destinations subject to the flow conservation condition given by equation (2.2) and imposing that there is no flow of information leaving the network, *i.e.*, $T(0) = 0$ and $T(L) = 0$. Thus the system of equations that model our problem in the one-dimensional case is given by:

$$\text{Min}\, N(0, L) = \int_0^L \eta(x)\,dx, \tag{2.3}$$

subject to the following constraints

$$\frac{dT(x)}{dx} = \rho(x) \quad \text{for all } x \in (0, L), \tag{2.4}$$

$$T(0) = 0 \quad \text{and} \quad T(L) = 0. \tag{2.5}$$

Notice that in the one-dimensional case, there is no minimization problem since by using the constraints (2.4) and (2.5), we obtain just one solution. As we will see, this will not

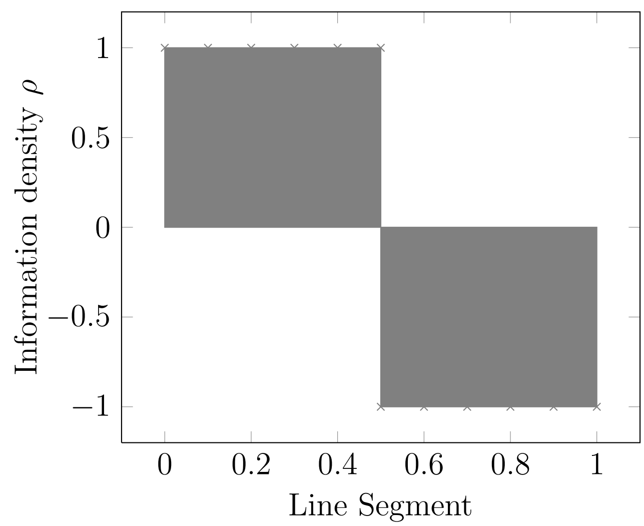


Figure 2.1: The information density function over the line segment [0 m, 10 m] in Example 2.1.1, given in the first half of the line segment [0 m, 5 m] by a uniform information density function generated by the sources of $\rho(x) = 1$ bps/m and in the second half [5 m, 10 m] by a uniform information density function received at the sensor destinations given by $\rho(x) = -1$ bps/m.

be the case for the two-dimensional case. Within the one-dimensional case context, we further assume that the proportion of sensor nodes $\eta(x)$ in a line segment of infinitesimal size $d\varepsilon$, centered at location x, needed as relay nodes, is proportional to the traffic flow of information that is passing through that region, *i.e.*, $\eta(x)\,d\varepsilon = \|T(x)\|^\alpha\,d\varepsilon$ where $\alpha > 0$ is a fixed number called *the relay-traffic constant*. Then the optimal placement of the relay nodes in the network will be given by $\eta^*(x) = \|T^*(x)\|^\alpha$, where the traffic flow function $T^*(x)$ is the optimal traffic flow function, given by the solution of the previous system of equations. Furthermore, the optimal total number of relay nodes $N^*(0, L)$ needed to support the optimal traffic flow function $T^*(x)$ in the network will be

$$N^*(\ell_0, \ell_1) = \int_{\ell_0}^{\ell_1} \eta(x)\,dx = \int_{\ell_0}^{\ell_1} \|T(x)\|^\alpha\,dx.$$

Let us see an example to illustrate the previous framework.

Example 2.1.1 *Suppose that we can divide the line segment* $[0, L]$ *in two parts:*

- *In the first half* $[0, L/2]$ *there will be a uniform information density function generated by the sensor sources, given by* $\rho(x) = 1$ *bps/m.*
- *In the second half* $[L/2, L]$ *there will be a uniform information density function received at the sensor destinations given by* $\rho(x) = -1$ *bps/m (see Figure 2.1).*

From the equations (2.4) *and* (2.5) *we obtain that the optimal traffic flow function will be given by*

$$T^*(x) = \begin{cases} x & bps/m \quad \textit{for all } x \in [0, L/2] \\ L - x & bps/m \quad \textit{for all } x \in [L/2, L] \end{cases}$$

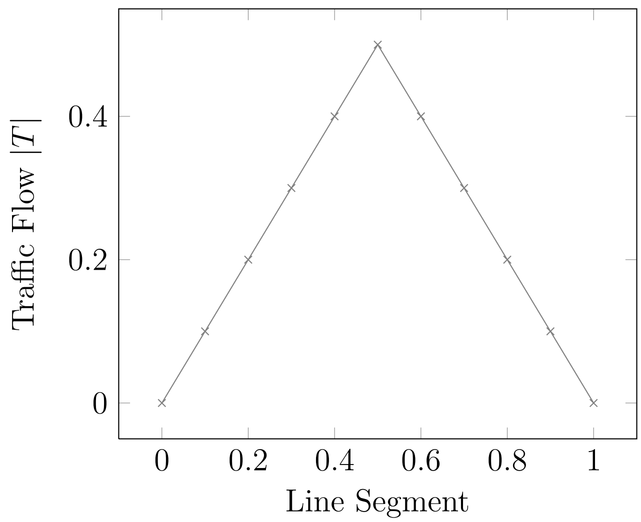


Figure 2.2: Optimal magnitude of the traffic flow T^* with positive direction in Example 2.1.1 where in the first half of the line segment $[0\text{ m}, 5\text{ m}]$ there is a uniform information density function generated by the sensor sources of $\rho(x) = 1$ bps/m and in the second half $[5\text{ m}, 10\text{ m}]$ there is a uniform information density function received at the sensor destinations given by $\rho(x) = -1$ bps/m.

with positive direction (see Figure 2.2). If we further assume that the relay-traffic constant $\alpha = 2$, then the optimal placement of the relay nodes needed to relay the information from the sources to the destinations on the network will be given by (see Figure 2.3)

$$\eta^*(x) = \begin{cases} x^2 & \textit{nodes} \quad \textit{for all } x \in [0, L/2] \\ (L-x)^2 & \textit{nodes} \quad \textit{for all } x \in [L/2, L] \end{cases}$$

The optimal total number of relay nodes $N^*(0, L)$ needed to support the optimal traffic flow $T^*(x)$ will be given by

$$N^*(L) = \int_0^{L/2} x^2\, dx + \int_{L/2}^{L} (L-x)^2\, dx = L^3/12.$$

From this example we obtained a closed-form expression for the total number of nodes needed to maintain the optimal traffic flow as a function of the length of the line segment for the one-dimensional case. Within this context the problem didn't required any minimization. This will not be the case for the two-dimensional case.

2.2 Fluid approximations: two-dimensional case

Consider a grid area network D in the two dimensional plane[1] $X_1 \times X_2$. Consider the continuous *information density function* $\rho(\mathbf{x})$, measured in bps/m^2, such that at locations

[1]We will denote with bold fonts the vectors and $\mathbf{x} = (x_1, x_2)$ will denote a location in the two dimensional plane $X_1 \times X_2$.

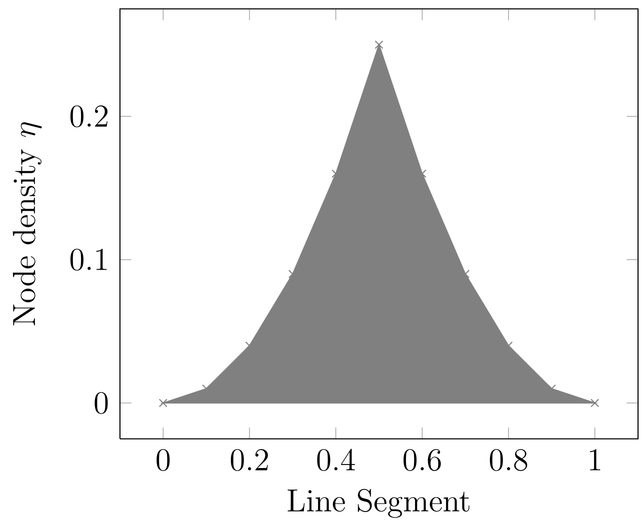


Figure 2.3: Optimal placement of the relay nodes η^* in Example 2.1.1 where in the first half of the line segment $[0 \text{ m}, 5 \text{ m}]$ there is a uniform information density function generated by the sensor sources of $\rho(x) = 1$ bps/m and in the second half $[5 \text{ m}, 10 \text{ m}]$ there is a uniform information density function received at the sensor destinations given by $\rho(x) = -1$ bps/m.

$\mathbf{x}$ where $\rho(\mathbf{x}) > 0$, there is a distributed data created by sources, such that the rate with which information is created in an infinitesimal area of size dA_ε, centered at location $\mathbf{x}$, is $\rho(\mathbf{x})\, dA_\varepsilon$. Similarly, at locations $\mathbf{x}$ where $\rho(\mathbf{x}) < 0$, there is a distributed data received at destinations, such that the rate with which information can be treated by an infinitesimal area of size dA_ε, centered at location $\mathbf{x}$, is equal to $-\rho(\mathbf{x})\, dA_\varepsilon$.

The total rate at which sensor destinations must process data is the same as the total rate which the data is created at the sensor sources, *i.e.*,

$$\int_{X\times Y} \rho(\mathbf{x})\, d\mathbf{x} = 0.$$

Consider the continuous *node density function* $\eta(\mathbf{x})$, measured in nodes/m^2, defined so that the number of relay nodes in an area of infinitesimal size dA_ε, centered at $\mathbf{x}$, is equal to $\eta(\mathbf{x})\, dA_\varepsilon$.

The total number of nodes on a region A, denoted by $N(A)$, is then given by

$$N(A) = \int_A \eta(\mathbf{x})\, d\mathbf{x}.$$

Consider the continuous *traffic flow function* $\mathbf{T}(\mathbf{x})$, measured in bps/m, such that its direction coincides with the direction of the flow of information at point $\mathbf{x}$, and[2] $\|\mathbf{T}(\mathbf{x})\|$ is the rate with which information rate crosses a linear segment perpendicular to $\mathbf{T}(\mathbf{x})$ centered on $\mathbf{x}$, *i.e.*, $\|\mathbf{T}(\mathbf{x})\|\, \varepsilon$ gives the total amount of traffic crossing a linear segment of infinitesimal length ε, centered at location $\mathbf{x}$, and placed vertically to $\mathbf{T}(\mathbf{x})$.

[2] The norm $\|\cdot\|$ is the Euclidean norm, *i.e.*, for a vector $\mathbf{x} = (x_1, x_2)$, its norm will be $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$.

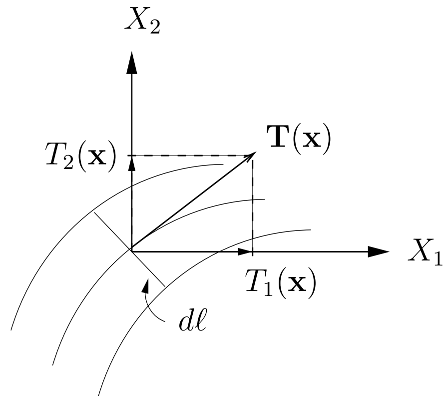


Figure 2.4: Flow of information $\mathbf{T}(\mathbf{x})$ through incremental line segment $d\ell$, decomposed in its horizontal component $T_1(\mathbf{x})$ (in the direction X_1) and its vertical component $T_2(\mathbf{x})$ (in the direction X_2).

Next we present the flow conservation condition (see *e.g.* [24, 29], for more details about this type of condition). For information to be conserved over a domain D of arbitrary shape on the $X \times Y$ plane, with smooth boundary S, it is necessary that the rate with which information is created in the area is equal to the rate with which information is leaving the area, *i.e.*,

$$\int_D \rho(\mathbf{x})\, dD = \oint_S [\mathbf{T} \cdot \mathbf{n}\,(\mathbf{x})]\, d\ell \tag{2.6}$$

The integral on the left-hand side is the surface integral of $\rho(\mathbf{x})$ over the domain D. The integral on the right-hand side is the path integral of the inner product $\mathbf{T} \cdot \mathbf{n}$ over the boundary S. The vector $\mathbf{n}(\mathbf{x})$ is the unit normal vector to S at the boundary point $\mathbf{x} \in S$ and pointing outwards. Then the function $\mathbf{T} \cdot \mathbf{n}\,(\mathbf{x})$, measured in bps/m^2, is equal at the rate with which information is leaving the domain D at the boundary point $\mathbf{x}$.

This holding for any (smooth) domain D, it follows that necessarily

$$\nabla \cdot \mathbf{T}(\mathbf{x}) := \frac{\partial T_1(\mathbf{x})}{\partial x_1} + \frac{\partial T_2(\mathbf{x})}{\partial x_2} = \rho(\mathbf{x}), \tag{2.7}$$

where "$\nabla \cdot$" is the divergence operator. Notice that equations (2.6) and (2.7) are the integral and differential versions of Gauss's law, respectively.

Thus the problem considered is to minimize the quantity of nodes $N(D)$ in the grid area network D needed to support the information created by the distribution of sources subject to the flow conservation condition, *i.e.*, our problem is given by the system of equations:

$$\text{Min} \quad N(D) \tag{2.8}$$

$$\text{subject to} \quad \nabla \cdot \mathbf{T} = \rho(\mathbf{x}). \tag{2.9}$$

Tassiulas and Toumpis prove in [48] that among all traffic flow functions that satisfy $\nabla \cdot \mathbf{T} = \rho$, the one that minimizes the number of nodes needed to support the throughput demands of the network, must be irrotational, *i.e.*,

$$\nabla \times \mathbf{T} = 0. \tag{2.10}$$

where "$\nabla \times$" is the curl operator.

Extension to multi-class traffic

The work on massively dense ad hoc networks considers a single class of traffic. In the Geometrical Optics approach it corresponds to the demand at location **A** from location **B**. In Electrostatics it corresponds to a set of origins and a set of destinations where traffic from any origin point could go to any destination point. The analogy to positive and negative charges in Electrostatics may limit the perspectives of multi-class problems where traffic from distinct origin sets has to be routed to distinct destination sets.

The model based on Geometrical Optics can directly be extended to include multiple classes as there are no elements in the model that suggest coupling between classes. This is due in particular to the fact that the cost density has been assumed to depend only on the density of the nodes and not on the density of the flows.

In contrast, the cost in the model based on Electrostatics is assumed to depend both on the location as well as on the local flow density. It thus models more complex interactions that would occur if we considered the case of m traffic classes. Extending the relation (2.7) to the multi-class case, we have traffic conservation at each point in space for each traffic class as expressed in the following:

$$\forall k \in \{1, \ldots, m\} \quad \nabla \cdot \mathbf{T}^k(\mathbf{x}) = \rho^k(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathcal{D}. \tag{2.11}$$

The function $\mathbf{T}^k$ is the flow distribution of class k and ρ^k corresponds to the distribution of the external origin and/or destinations.

Let $\mathbf{T}(\mathbf{x})$ be the total flow vector at point $\mathbf{x} \in \mathcal{D}$. It is a vector of dimension m, and each one of the m-entries is a two dimensional flow. A generic multi-class optimization problem would then be: minimize Z over the flow distributions $\{\mathbf{T}^k\}$

$$Z = \int_{\mathcal{D}} c(\mathbf{x}, \mathbf{T}(\mathbf{x}))\, d\mathbf{x} \quad \text{subject to} \quad \nabla \cdot \mathbf{T}^k(\mathbf{x}) = \rho^k(\mathbf{x}),\ k \in \{1, ..., m\}, \qquad \forall \mathbf{x} \in \mathcal{D}. \tag{2.12}$$

Chapter 3

Directional Antennas: User- and System-Optimization

So far we have adopted a general framework under which the information is conserved. Through this chapter we make more specific assumptions on the cost function. The particular cost function introduced here can be called an ℓ^1-norm cost model, in which the transmission cost from a location to another is the sum of the horizontal and vertical transmission cost components. This assumption is justified in the case when the information can be transmitted only horizontally or vertically (so that even a continuous diagonal curve is understood as the limit of many horizontal and vertical transmissions into smaller hop distances). In road traffic engineering, this corresponds to a Manhattan-like network, see *e.g.* [29]. In the context of ad hoc networks this would correspond to directional antennas (with either horizontal or vertical direction). In this type of networks we are able to provide a simple characterization for the minimum cost paths with a particular cost structure.

3.1 Basics on directional antennas

Consider a wireless ad hoc network where a large number of nodes are placed deterministically in the grid area network D. For energy efficiency reasons, we assume that each node is equipped with one or two directional antennas, allowing transmissions at each hop to be directed either from North-to-South or from West-to-East following the notation convention used in the work of Dafermos [29]. We extend the work of Dafermos [29] in road traffic engineering to the multi-class problem into the directional antennas framework. We consider m classes of transmissions in each of these directions: $T_1^k \geq 0$, $k \in \{1, ..., m\}$ (West-to-East, taken as the positive direction of the axis x_1), $T_2^k \geq 0$, $k \in \{1, ..., m\}$ (North-to-South, taken as the positive direction of the axis x_2). In a massively dense ad hoc network, or continuous approximation of the network, a curved path can be viewed as a limit of a path with many such hops as the hop distance tends to zero.

Some assumptions on the cost function:

- *User cost*
  - We allow the cost for a horizontal transmission (West-to-East, or equivalently, in the direction of the axis x_1) to be different than the cost for a vertical transmission (North-to-South, or equivalently, in the direction of the axis x_2).
    It is assumed that a packet traveling in the direction of the axis x_1 incurs in a *transmission cost* of g_1, and equivalently, traveling in the direction of the axis x_2 incurs in a *transmission cost* of g_2.
  - *Congestion dependent cost:* Notice that the transmission costs g_1 and g_2 depend on the location $\mathbf{x}$ and the traffic flow $\mathbf{T}(\mathbf{x})$ that is flowing through that location, *i.e.*,
    $$g_1 = g_1(\mathbf{x}, \mathbf{T}(\mathbf{x})) \quad \text{and} \quad g_2 = g_2(\mathbf{x}, \mathbf{T}(\mathbf{x})).$$
  - We consider a *vector transmission cost* $\mathbf{g} := (g_1, g_2)$. Notice that as each of its components, such vector transmission cost also depends upon the location $\mathbf{x}$ and the traffic flow $\mathbf{T}(\mathbf{x})$ flowing through that location, *i.e.*, $\mathbf{g} = \mathbf{g}(\mathbf{x}, \mathbf{T}(\mathbf{x}))$.
  - The *local transmission cost* g is given by the inner product between the vector transmission cost and the flow of information, *i.e.*,
    $$g(\mathbf{x}, \mathbf{T}(\mathbf{x})) = \mathbf{g}(\mathbf{x}, \mathbf{T}(\mathbf{x})) \cdot \mathbf{T}(\mathbf{x}) = g_1 T_1 + g_2 T_2,$$
    and it corresponds to the sum of the transmission costs multiplied by the quantity of flow in each direction.
  - The local transmission cost $g(\mathbf{x}, \mathbf{T}(\mathbf{x}))$ is assumed to be non-negative, monotone increasing in each component of $\mathbf{T}$ (T_1 and T_2 in our 2-dimensional case).
- *System cost*
  - The *system transmission cost* is the integral of the local transmission cost over the network, *i.e.*,
    $$\int_{\mathcal{D}} g(\mathbf{x}, \mathbf{T}(\mathbf{x}))\, d\mathbf{x}.$$

The **boundary conditions** are determined by the options that users have in selecting their origin and/or destinations. Notice that, with abuse of terminology, we use the generic term "users" to denote packets or data in the network. Another abuse of notation, is that the decision about which destination to transmit each packet is not done by the packets or data itself but by the network configuration in which each node decides to which destination to transmit in order to minimize the cost function. Examples of boundary conditions are:

- *Assignment problem*: users of the network have predetermined origin and destinations and are free to choose their travel paths.
- *Combined distribution and assignment problem*: users of the network have predetermined origins and are free to choose their destinations (within a certain destination region) as well as their travel paths.

- *Combined generation, distributions and assignment problem*: users are free to choose their origins, their destinations, as well as their travel paths.

One important application of wireless ad hoc networks is "area monitoring of sensor nodes" where there are one or several data aggregation centers of the phenomenon being monitored. The problem when there is only one data aggregation center can be seen as an assignment problem where the origins are the sensor nodes location where the phenomenon is being monitored and the destination is the unique data aggregation center. When there are several data aggregation centers it can be seen as a combined distribution and assignment problem where the packet or data can be transmitted to any of the data aggregation centers. In this chapter we are interested in these types of boundary conditions.

The problem formulation is again to minimize Z as defined in (2.12). The natural choice of functional spaces to make the problem precise, and to take advantage of the available theory developed in the PDE (Partial Differential Equations) community, is to work in the Sobolev space $H^1(\mathcal{D})$. In order to define this space, we need to introduce two other concepts:

- We define $L^2(\mathcal{D})$ as the space of functions that are square-integrable in $\mathcal{D}$, *i.e.*,

$$L^2(\mathcal{D}) = \left\{ f \text{ measurable, such that } \int_{\mathcal{D}} \|f(\mathbf{x})\|^2 \, d\mathbf{x} < +\infty \right\}.$$

- Take f to be a scalar function, define its *weak gradient* as a vector function g such that, for any smooth vector function φ with compact support in D,

$$\int_D f \; \nabla \cdot \varphi = - \int_D < g, \varphi > .$$

We define $H^1(\mathcal{D})$ as the space of functions that are square integrable, and with weak gradient square-integrable. Then our objective is to find the optimal traffic flow functions T_1^k, $k \in \{1, \dots, m\}$ and T_2^k, $k \in \{1, \dots, m\}$ in $H^1(\mathcal{D})$, for an information density function ρ in $L^2(\mathcal{D})$.

3.2 Karush-Kuhn-Tucker conditions

The Karush-Kuhn-Tucker conditions (KKT conditions), which we introduce below, are necessary conditions for a solution to be optimal in nonlinear programming. It is a generalization of the method of Lagrange multipliers to inequality constraints. We recall that the method of Lagrange multipliers provides a strategy for finding the minimum of a function subject to constraints and it is based on introducing new variables called Lagrange multipliers and study a new function called Lagrange function. In the method of Lagrange multipliers, if a point is optimal for the minimization problem then there exist Lagrange multipliers such that the point is stationary for the Lagrange function. Notice however that not all stationary points for the Lagrange function yield a solution to the original problem. Thus, this method yields necessary conditions for optimality. However, it helps us to restrict our candidates to be optimal. And under some extra conditions there are some sufficient conditions on the results.

In the problem considered here, the traffic flows in each direction T_1^k and T_2^k are functionals (maps from the vector space to the scalar space). Then we have to consider variational inequalities, *i.e.*, inequalities involving a functional which have to be solved for all the values of the vector space.

A key application of these KKT conditions will be introduced in Section 3.5 where we study the user-optimization (non-cooperative) problem for each packet sent through the network and show that the solution must satisfy some variational inequalities. We then show that these inequalities can be interpreted as the Karush-Kuhn-Tucker conditions that we introduce in this chapter, applied to some *transformed* cost (called "potential" and introduced by Beckmann [7]). This will allow us to propose a method for solving the user-optimization (non-cooperative) problem.

We begin by recalling Green's Theorem which has proved to be useful for many physical phenomena. We will make extensive use of this theorem in this section as well as in Section 3.5.

Theorem 3.2.1 (Green's Theorem) *Let $\mathcal{D}$ be a region of the space, and let S be its piecewise-smooth boundary, for all $x \in S$, $\mathbf{n}$ a unit outward normal to $\mathcal{D}$. Consider the scalar function u and a continuously differentiable vector function $\mathbf{v}$, then*

$$\int_{\mathcal{D}} u \nabla \cdot \mathbf{v} dx = \int_S u < \mathbf{v}, \mathbf{n} > d\ell - \int_{\mathcal{D}} < \mathbf{v}, \nabla u > dx.$$

Using this theorem and the Karush-Kuhn-Tucker conditions we are able to prove a result that provides a characterization of the optimal solution for some special cases as we will see in Section 3.4.

Theorem 3.2.2 *Define the Lagrange function as*

$$L^{\zeta}(\mathbf{T}) := \int_{\mathcal{D}} \ell^{\zeta}(\mathbf{x}, \mathbf{T})\, d\mathbf{x} \quad \textit{with} \quad \ell^{\zeta}(\mathbf{x}, \mathbf{T}) := g(\mathbf{x}, \mathbf{T}) - \sum_{j=1}^{m} \zeta^j(\mathbf{x}) \Big[\nabla \cdot \mathbf{T}^j(\mathbf{x}) - \rho^j(\mathbf{x}) \Big]$$

where $\zeta^j(\mathbf{x}) \in L^2(\mathcal{D})$ are called Lagrange multipliers.

For a vector field $\mathbf{T}(\cdot)$ with positive components satisfying (2.11), a necessary and sufficient condition for minimizing the cost (2.12) is that the Lagrangian be minimized over all vector fields with positive components, or equivalently, that equations

$$\frac{\partial g(\mathbf{x}, \mathbf{T})}{\partial T_i^j} + \frac{\partial \zeta^j(\mathbf{x})}{\partial x_i} = 0 \quad \textit{if} \quad T_i^j(\mathbf{x}) > 0, \tag{3.1a}$$

$$\frac{\partial g(\mathbf{x}, \mathbf{T})}{\partial T_i^j} + \frac{\partial \zeta^j(\mathbf{x})}{\partial x_i} \geq 0 \quad \textit{if} \quad T_i^j(\mathbf{x}) = 0. \tag{3.1b}$$

be satisfied.

Proof.- The criterion is convex, and the constraint (2.11) is affine. Therefore the Karush-Kuhn-Tucker theorem holds, stating that the Lagrangian is minimum at the optimum. A variation $\delta\mathbf{T}(\cdot)$ will be admissible if $\mathbf{T}(\mathbf{x})+\delta\mathbf{T}(\mathbf{x})\geq 0$ for all $\mathbf{x}$, hence in particular, for all $\mathbf{x}$ such that $T_i^j(\mathbf{x})=0$ and $\delta T_i^j(\mathbf{x})\geq 0$.

As we are working with functionals, we need a generalization of the concept of directional derivative used in differential calculus. The Gâteaux differential $DF(u,d)$ of functional F at u in the direction d is defined as

$$DF(u,d)=\lim_{t\to 0}\frac{F(u+td)-F(u)}{t}=\frac{d}{dt}F(u+td)\Big|_{t=0}.$$

if the limit exists. If the limit exists for all d, one says that F is Gâteaux differentiable at u.

Let $\mathrm{D}L^\zeta$ denote the Gâteaux derivative of functional L^ζ with respect to $\mathbf{T}(\cdot)$. First order condition for local minimum reads:

$$\text{For all } \delta\mathbf{T} \text{ admissible}, \mathrm{D}L^\zeta\cdot\delta\mathbf{T}\geq 0\,,$$

therefore here

$$\int_{\mathcal{D}}\sum_j\langle\nabla_{\mathbf{T}^j}g(\mathbf{x},\mathbf{T}(\mathbf{x})),\delta\mathbf{T}^j(\mathbf{x})\rangle\,d\mathbf{x}-\int_{\mathcal{D}}\sum_j\zeta^j(\mathbf{x})\nabla\cdot\delta\mathbf{T}^j(\mathbf{x})\,d\mathbf{x}\geq 0.$$

Integrating by parts using Green's Theorem, this is equivalent to

$$\int_{\mathcal{D}}\sum_j\left[\langle\nabla_{\mathbf{T}^j}g,\delta\mathbf{T}^j\rangle+\langle\nabla_{\mathbf{x}}\zeta^j,\delta\mathbf{T}^j\rangle\right]d\mathbf{x}-\int_{\partial\mathcal{D}}\sum_j\zeta^j\langle\delta\mathbf{T}^j,\mathbf{n}\rangle\,d\ell\geq 0\,.$$

We may choose all the components $\delta\mathbf{T}^k=0$ except $\delta\mathbf{T}^j$, and choose $\delta\mathbf{T}^j$ in $(H_0^1(\mathcal{D}))^2$, *i.e.*, functions in $H^1(\mathcal{D})$ such that their boundary integral be zero. This is always feasible and admissible. Then the last term above vanishes, and it is a classical fact that the inequality implies (3.1a)-(3.1b) for $i=1,2$. Placing this back in Euler's inequality, and using a $\delta\mathbf{T}^j$ non zero on the boundary, it follows that necessarily[1] $\zeta^j(\mathbf{x})=0$ at any $\mathbf{x}$ of the boundary S where $T(\mathbf{x})>0$. As we shall see, this conditions provides the boundary condition to recover the Lagrange multipliers ζ^j from equation (2.11). Equations (3.1a)-(3.1b) are already stated in [29] for the single class case. However, as Dafermos states explicitly, its rigorous derivation is not available there. ∎

Consider the following special cases that we shall need later. We assume a single traffic class, but this could easily be extended to several traffic classes. Let

$$g(\mathbf{x},\mathbf{T}(\mathbf{x}))=\sum_{i=1,2}g_i(\mathbf{x},\mathbf{T}(\mathbf{x}))T_i(\mathbf{x}).$$

- Monomial cost per packet:

$$g_i(\mathbf{x},\mathbf{T}(\mathbf{x}))=k_i(\mathbf{x})\Big(T_i(\mathbf{x})\Big)^\beta \tag{3.2}$$

[1] This is a complementary slackness condition on the boundary.

for some $\beta > 1$. Then (3.1a)-(3.1b) simplify to

$$(\beta+1)k_i(\mathbf{x})\left(T_i(\mathbf{x})\right)^{\beta} + \frac{\partial\zeta(\mathbf{x})}{\partial x_i} = 0 \quad \text{if } T_i(\mathbf{x}) > 0, \tag{3.3a}$$

$$(\beta+1)k_i(\mathbf{x})\left(T_i(\mathbf{x})\right)^{\beta} + \frac{\partial\zeta(\mathbf{x})}{\partial x_i} \geq 0 \quad \text{if } T_i(\mathbf{x}) = 0. \tag{3.3b}$$

In that case, recovery of ζ to complete the process is difficult, at best. Things are much simpler in the next case.

- Affine cost per packet:

$$g_i(\mathbf{x}, \mathbf{T}(\mathbf{x})) = \frac{1}{2}k_i(\mathbf{x})T_i(\mathbf{x}) + h_i(\mathbf{x}). \tag{3.4}$$

Then, equations (3.1a)-(3.1b) simplify to

$$k_i(\mathbf{x})T_i(\mathbf{x}) + h_i(\mathbf{x}) + \frac{\partial\zeta(\mathbf{x})}{\partial x_i} = 0 \quad \text{if } T_i(\mathbf{x}) > 0,$$

$$k_i(\mathbf{x})T_i(\mathbf{x}) + h_i(\mathbf{x}) + \frac{\partial\zeta(\mathbf{x})}{\partial x_i} \geq 0 \quad \text{if } T_i(\mathbf{x}) = 0.$$

Assume that the $k_i(\cdot)$'s are positive everywhere and bounded away from 0. For simplicity, let $a_i = 1/k_i$, and b be the vector with coordinates $b_i = h_i/k_i$, all assumed to be square integrable. Assume that there exists a solution where $T(\mathbf{x}) > 0$ for all $\mathbf{x}$. Then

$$T_i(\mathbf{x}) = -\left(a_i(\mathbf{x})\frac{\partial\zeta(\mathbf{x})}{\partial x_i} + b_i(\mathbf{x})\right).$$

As a consequence, from (2.11) and the above remark, we get that $\zeta(\cdot)$ is to be found as the solution in $H_0^1(\mathcal{D})$ of the elliptic equation (an equality in $H^{-1}(\mathcal{D})$)

$$\sum_i \frac{\partial}{\partial x_i}\left(a_i(\mathbf{x})\frac{\partial\zeta(\mathbf{x})}{\partial x_i}\right) + \nabla\cdot b(\mathbf{x}) + \rho(\mathbf{x}) = 0\,.$$

This is a well behaved Dirichlet problem, known to have a unique solution[2] in $H_0^1(\mathcal{D})$, furthermore easy to compute numerically.

3.3 User-optimization and congestion independent costs

In this section, we extend the shortest path approach for optimization that has already appeared using Geometrical Optics tools [20]. We present a general optimization framework for handling shortest path or minimum cost paths problems. We assume that the local transmission cost depends on the direction of the flow but not on its size, *i.e.*, it is a congestion

[2]There are many cases where the equilibrium may not be unique (see *e.g.* [53]).

independent cost as explained in subsection 1.2.1. The cost is $c_1(\mathbf{x})$ for a flow that is locally horizontal and is $c_2(\mathbf{x})$ for a flow that is locally vertical. As previously stated, we assume in this section that c_1 and c_2 do not depend on $\mathbf{T}$. The cost incurred by a packet transmitted along a path p is given by the line integral

$$\mathbf{c}_p = \int_p \mathbf{c} \cdot \, d\mathbf{x}. \tag{3.6}$$

Let $V^j(\mathbf{x})$ be the minimum cost to go from a point $\mathbf{x}$ to a set B^j, $j = 1, ..., m$. Then

$$V^j(\mathbf{x}) = \min \left(c_1(\mathbf{x})\, dx_1 + V^j(x_1 + \, dx_1, x_2), c_2(\mathbf{x})\, dx_2 + V^j(x_1, x_2 + \, dx_2) \right). \tag{3.7}$$

This can be written as the Hamilton-Jacobi-Bellman (HJB) equation:

$$0 = \min \left(c_1(\mathbf{x}) + \frac{\partial V^j(\mathbf{x})}{\partial x_1}, c_2(\mathbf{x}) + \frac{\partial V^j(\mathbf{x})}{\partial x_2} \right), \qquad \forall \mathbf{x} \in B^j \, , \; V^j(\mathbf{x}) = 0 \, . \tag{3.8}$$

If V^j is differentiable then, under suitable conditions, it is the unique solution of (3.8). In the case that V^j is not everywhere differentiable then, under suitable conditions, it is the unique viscosity solution of (3.8) (see [35, 36]).

There are many numerical approaches for solving the Hamilton-Jacobi-Bellman (HJB) equation. One can discretize the HJB equation and obtain a discrete dynamic programming for which efficient solution methods exist. If one repeats this for various discretization steps, then we know that the solution of the discrete problem converges to the viscosity solution of the original problem (under suitable conditions) as the step size converges to zero [35]. In next section we will see nice characterization of the paths by considering some particular transmission cost structure.

3.4 Characterization of minimum cost paths

We consider now our directional antenna model in a given rectangular area R, defined by the simple closed curve $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$ (see Fig. 3.1). We study the case where transmissions can go from North-to-South or from West-to-East. We obtain below *optimal paths* defined as paths that achieve the minimum packet transmission cost defined by (3.6). We shall study two problems:

- *Point to point optimal path:* we seek the minimum cost path between two points. This corresponds to the combined distribution and assignment problem when the destination region is a point or it can be considered as the assignment problem as explained in Section 3.1.

- *Point to boundary optimal path:* we seek the minimum cost path on a given region that starts at a given point and is allowed to end at any point on the boundaries. This corresponds to the case of combined distribution and assignment problem.

Another formulation of Green's Theorem, stated previously as Theorem 3.2.1, gives us a characterization of the optimal paths for those two problems.

Theorem 3.4.1 (Green's Theorem: alternative version) *Let $\mathcal{D}$ be a region of the space, and let S be its piecewise-smooth boundary. Suppose that P and Q are continuously differentiable functions in $\mathcal{D}$. Then*

$$\oint_S Pdx + Qdy = \int_{\mathcal{D}} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy.$$

Recall that the cost is composed of a horizontal and a vertical component (these are $c_1(\mathbf{x})$ and $c_2(\mathbf{x})$ respectively), which do not depend on the traffic flow.

Consider the function

$$U(\mathbf{x}) = \frac{\partial c_2}{\partial x_1}(\mathbf{x}) - \frac{\partial c_1}{\partial x_2}(\mathbf{x}).$$

It will turn out that the structure of the minimum cost path depends on the costs through the sign of the function U. Now, if the function $\mathbf{c}$ is continuously differentiable then U is a continuous function. This motivates us to study cases in which U has the same sign everywhere (see Fig. 3.1), or in which there are two regions in the rectangle R, one with $U > 0$ and one with $U < 0$, separated by a curve ℓ on which $U = 0$ (see *e.g.* Fig. 3.2). We shall

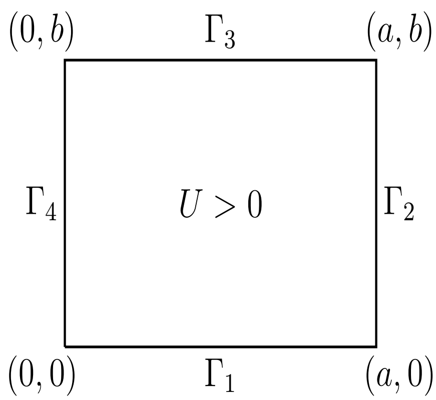


Figure 3.1: The rectangle R defined by the boundaries $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$. The case where $U > 0$.

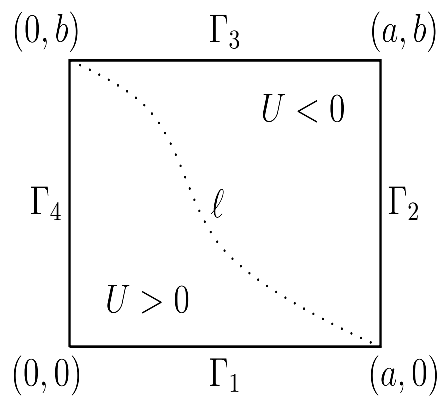


Figure 3.2: The case of two regions separated by a curve. Case 1.

assume throughout that the function $\mathbf{c}$ is continuously differentiable, and that, if non-empty, the set of points inside the domain where the function U is zero, *i.e.*, $\ell = \{\mathbf{x} : U(\mathbf{x}) = 0\}$, is a smooth line. (This is true, *e.g.*, if $\mathbf{c}$ is a smooth function and $\nabla U \neq 0$ on ℓ.)

The function U has the same sign over the whole region

Theorem 3.4.2 *(Point to point optimal path) Suppose that an origin point $\mathbf{x}^o = (x_1^o, x_2^o)$ wants to send a packet to a destination point $\mathbf{x}^d = (x_1^d, x_2^d)$ and both points are in the interior of a rectangle R.*

i. If the function U is positive almost everywhere in the interior rectangle R_{od} defined by both points (see Fig. 3.3(a)), then the optimal path $\gamma_{\rm opt}$ is given by a horizontal straight line γ_H and then a vertical straight line γ_V (see Fig. 3.3(a)).

More precisely, $\gamma_{\rm opt} = \gamma_H \cup \gamma_V$ where

$$\gamma_H = \{(x_1, x_2) \textit{ such that } x_1^o \leq x_1 \leq x_1^d, x_2 = x_2^o\},$$
$$\gamma_V = \{(x_1, x_2) \textit{ such that } x_1 = x_1^d, x_2^o \leq x_2 \leq x_2^d\}.$$

ii. If the function U is negative almost everywhere in the interior rectangle R_{od} then there is an optimal path $\gamma_{\rm opt}$ given by a vertical straight line γ_V and then a horizontal straight line γ_H (see Fig. 3.3(b)).

More precisely, $\gamma_{\rm opt} = \gamma_V \cup \gamma_H$ where

$$\gamma_V = \{(x_1, x_2) \textit{ such that } x_1 = x_1^o, x_2^o \leq x_2 \leq x_2^d\},$$
$$\gamma_H = \{(x_1, x_2) \textit{ such that } x_1^o \leq x_1 \leq x_1^d, x_2 = x_2^d\}.$$

iii. In both cases, γ^{opt} is unique almost surely (i.e., the area between γ^{opt} and any other optimal path is zero).

Proof.- Consider an arbitrary path[3] γ_C joining $\mathbf{x}^o$ to $\mathbf{x}^d$, and assume that the area between $\gamma^{\rm opt}$ and γ_C is nonzero. We call γ_C the comparison path (see Fig. 3.3(a) for the case $U > 0$ and Fig. 3.3(b) for the case $U < 0$).

(i) Showing that the cost over path $\gamma_{\rm opt}$ is optimal is equivalent to showing that the integral of the cost over the closed path $\mathcal{P}$ is negative. Hereby $\mathcal{P}$ is given by following $\gamma^{\rm opt}$ from the origin point $\mathbf{x}^o$ to the destination $\mathbf{x}^d$, and then returning from the destination $\mathbf{x}^d$ to the origin point $\mathbf{x}^o$ by moving along the comparison path γ_C in the reverse direction. This closed path is written as $\mathcal{P} = \gamma_{\rm opt} \cup \gamma_C^-$ and A denotes the bounded area described by $\mathcal{P}$. Using Green's Theorem we obtain

$$\oint_{\mathcal{P}} \mathbf{c} \cdot \mathbf{dx} = -\int_A U(\mathbf{x}) dS$$

which is strictly negative since $U > 0$ almost everywhere on the interior rectangle R_{od}. Decomposing the left integral, this concludes the proof of (i), and establishes at the same time the corresponding statement on uniqueness in (iii).
(ii) is obtained similarly. ■

Theorem 3.4.3 *(Point to boundary optimal path)*
Consider the problem of finding an optimal path from a point $\mathbf{x}^o$ in the rectangle R to the boundary $\Gamma_1 \cup \Gamma_2$.

[3] Respecting that each sub-path can be decomposed in sums of paths either from North to South or from West to East (or is a limit of such paths).

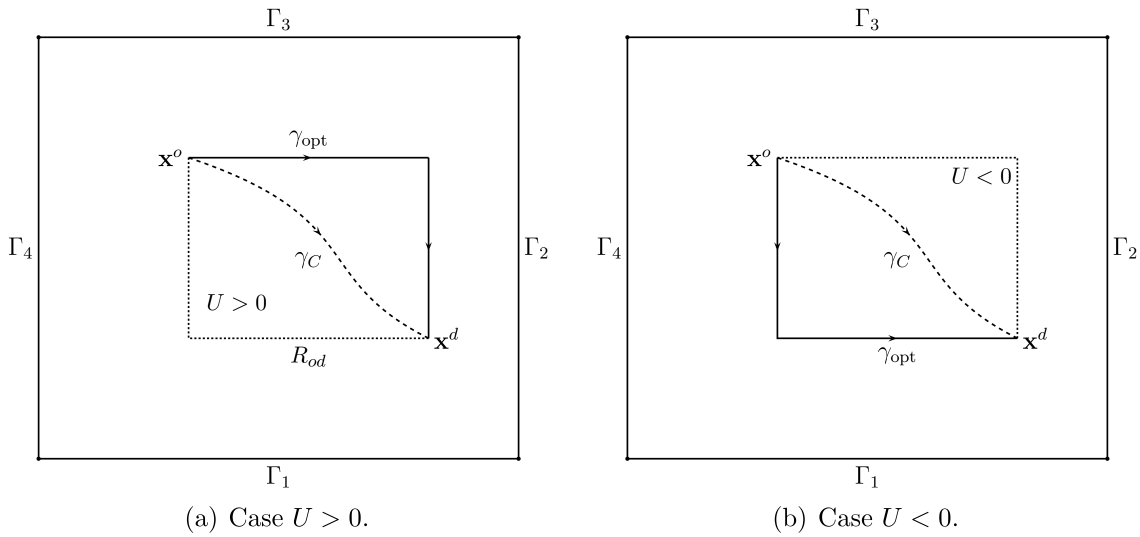


Figure 3.3: Optimal paths (a) when $U > 0$ and (b) when $U < 0$ in the interior rectangle defined by the origin point $\mathbf{x}^o$ and the destination point $\mathbf{x}^d$.

i. If the function U is almost everywhere negative inside the rectangle R and the cost on the boundary Γ_1 is non-negative and on boundary Γ_2 is non-positive, then the optimal path is the straight vertical line (see Fig. 3.4).

ii. If the function $U(\mathbf{x})$ is almost everywhere positive inside the rectangle R and the cost on the boundary Γ_1 is non-positive and on boundary Γ_2 is non-negative. Then the optimal path is the straight horizontal line (see Fig. 3.5).

Proof.-

(i) Denote by γ_{opt} the straight vertical path joining $\mathbf{x}^o$ to the boundary Γ_1. Consider another arbitrary valid path γ_C joining $\mathbf{x}^o$ to any point $\mathbf{x}^*$ on the boundary $\Gamma_1 \cup \Gamma_2$, and assume that the area between γ_{opt} and γ_C is nonzero. We call γ_C the comparison path.

Assume first that $\mathbf{x}^*$ is on the boundary Γ_2. Denote $\mathbf{x}^d$ the South-East corner of the rectangle R, *i.e.*, $\mathbf{x}^d := \Gamma_1 \cap \Gamma_2$. Then by Theorem 3.4.2(ii), the cost to go from $\mathbf{x}^o$ to $\mathbf{x}^d$ is smaller when using γ_{opt} and then continuing eastwards (along Γ_1) than when using the comparison path γ_C and then southwards (along Γ_2). Due to our assumptions on the costs over the boundaries, this implies that the cost along the straight vertical path γ_{opt} is smaller than along the comparison path γ_C.

Next consider the case where $\mathbf{x}^*$ is on the boundary Γ_1. Denote by η the section of the boundary Γ_1 that joins $\gamma_{\text{opt}} \cap \Gamma_1$ with $\mathbf{x}^*$ (see Figure 3.4). Then again, by Theorem 3.4.2 (ii), the cost to go from $\mathbf{x}^o$ to $\mathbf{x}^*$ is smaller when using γ_{opt} and then continuing eastwards (along Γ_1) than when using the comparison path γ_C. Due to our assumptions that the cost on Γ_1 is non-negative, this implies that the cost along γ_V is smaller than along γ_C.

(ii) is obtained similarly. ∎

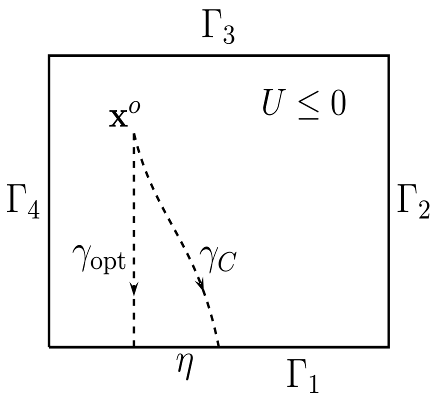


Figure 3.4: Theorem 3.4.3 (i)

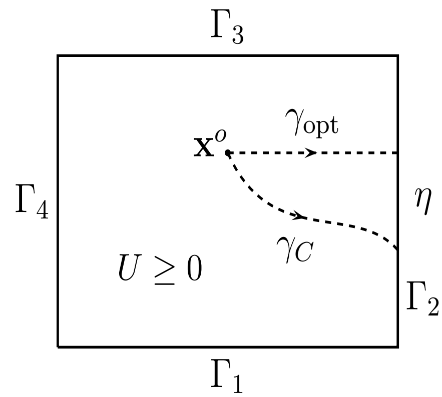


Figure 3.5: Theorem 3.4.3 (ii)

The function U changes sign within the region

Consider the region on the space $\ell := \{\mathbf{x} \in R \text{ such that } U(\mathbf{x}) = 0\}$. Let us consider the case when ℓ is a valid path in the rectangular area, such that it starts at the North-West corner (the intersection of the boundaries $\Gamma_3 \cap \Gamma_4$) and finishes at the South-East corner (the intersection of the boundaries $\Gamma_1 \cap \Gamma_2$). Then the space is divided in two areas, and as the function U is continuous we have the following cases:

1. $U(\mathbf{x})$ is negative in the upper area and positive in the lower area (see Fig. 3.2).

2. $U(\mathbf{x})$ is positive in the upper area and negative in the lower area.

The two other cases where the sign of U is the same over R are contained in what we solved in the previous subsection 3.4.

Case 1: The function $U(\mathbf{x})$ is negative in the upper area and positive in the lower area.

We shall show that in this case, ℓ is an attractor, in the sense that the optimal path reaches the line ℓ with the minimal possible distance and then continues along this line until it reaches the destination.

Proposition 3.4.1 *Assume that the origin point $\mathbf{x}^o$ and the destination point $\mathbf{x}^d$ are both on ℓ. Then the path p_ℓ that follows ℓ from the origin point $\mathbf{x}^o$ to the destination point $\mathbf{x}^d$ is optimal.*

Proof.- Consider a comparison path γ_C that coincides with ℓ only in the origin $\mathbf{x}^o$ and destination $\mathbf{x}^d$ points. First assume that the comparison path γ_C is entirely in the upper (*i.e.*, northern) part and call A the area between γ_C and p_ℓ. Define $\mathcal{P}$ to be the closed path that follows p_ℓ from $\mathbf{x}^o$ to $\mathbf{x}^d$ and then returns along γ_C.

The integral $\int_A U(\mathbf{x})\, d\mathbf{x}$ is negative by assumption. By Green's Theorem, it is equal to $\oint_{\mathcal{P}} \mathbf{c} \cdot d\mathbf{x}$. This implies that the cost along p_ℓ is strictly smaller than along γ_C.

A similar argument holds for the case that γ_C is below p_ℓ.

A path between $\mathbf{x}^o$ and $\mathbf{x}^d$ may have several intersections with ℓ. Between each pair of consecutive intersections of ℓ, the sub-path has a cost larger than that obtained by following ℓ between these points (this follows from the previous steps of the proof). We conclude that p_ℓ is indeed optimal. ∎

Proposition 3.4.2 *Let an origin point $\mathbf{x}^o$ send packets to a destination point $\mathbf{x}^D$.*

i. Assume both points are in the upper region. Denote by γ_1 the two segments path given by Theorem 3.4.2 (ii). Then the optimal curve γ_{opt} is obtained as the maximum between ℓ and γ_1[4].

ii. Let both points be in the lower region. Denote by γ_2 the two segments path given in Theorem 3.4.2 (i). Then the optimal curve γ_{opt} is obtained as the minimum between ℓ and γ_2.

Proof.-

(i) A straightforward adaptation of the proof of the previous proposition implies that the path in the statement of the proposition is optimal among all those restricted to the upper region. Consider now a path γ_C that is not restricted to the upper region. Then $\ell \cap \gamma_C$ contains two distinct points such that γ_C is strictly lower than ℓ between these points. Applying Proposition 3.4.1, we then see that the cost of γ_C can be strictly improved by following ℓ between these points instead of following γ_C there. This concludes (i).

(ii) Proved similarly. ∎

Proposition 3.4.3 *Let a point $\mathbf{x}^o$ send packets to a point $\mathbf{x}^d$.*

i. Assume the origin is in the upper region and the destination in the lower one. Then the optimal path has three segments;

1. It goes straight vertically from $\mathbf{x}^o$ to ℓ,

2. Continues as long as possible along ℓ, i.e., until it reaches the first coordinate of the destination,

3. At that point it goes straight vertically from ℓ to $\mathbf{x}^d$.

ii. Assume the origin is in the lower region and the destination in the upper one. Then the optimal path has three segments;

1. It goes straight horizontally from $\mathbf{x}^o$ to ℓ,

2. Continues as long as possible along ℓ, i.e., until it reaches the second coordinate of the destination,

[4] By the maximum we mean the following: If γ_1 does not intersect ℓ, then $\gamma_{\text{opt}} = \gamma_1$. If it intersects ℓ, then γ_{opt} agrees with γ_1 over the path segments where γ_1 is in the upper region and otherwise agrees with ℓ. The minimum is defined similarly.

3. At that point it goes straight horizontally from ℓ to $\mathbf{x}^d$.

Proof.- The proofs of (i) and of (ii) are the same. Consider an alternative route γ_C. Let $\tilde{\mathbf{x}}$ be some point in $\gamma_C \cap \ell$. The proof now follows by applying the previous proposition to obtain first the optimal path between the origin and $\tilde{\mathbf{x}}$ and second, the optimal path between $\tilde{\mathbf{x}}$ and the destination. ■

Case 2: The function U is positive in the upper area and negative in the lower area.

This case turns out to be more complex than the previous one. The curve M has some obvious repelling properties which we state next, but they are not as general as the attractor properties that we had in the previous case.

Proposition 3.4.4 *Assume that both origin and destination are in the same region. Then the paths that were optimal in Theorem 3.4.2 are optimal here as well, if we restrict ourselves to paths that remain in the same region.*

Proof.- Given that the origin and destination are in a region we may change the cost over the other region so that it has the same sign over all the region R. This does not influence the cost of path restricted to the region of the origin-destination pair. With this transformation we are in the scenario of Theorem 3.4.2 which we can then apply. ■

Discussion.- Note that the (sub)optimal policies obtained in Proposition 3.4.4 indeed look like being repelled from ℓ; their two segments trajectory guarantees to go from the origin to the destination as far as possible from ℓ.

We note that unlike the attracting structure that we obtained in Case 1, one cannot extend the repelling structure to the case where the paths are allowed to traverse from one region to another.

3.5 User-optimization and congestion dependent cost

We now go beyond the approach of the previous sections by allowing the cost to depend on congestion. In this setting, minimum cost paths can be a system objective as we shall motivate below. But it can also be the result of decentralized decision making by many "infinitesimally small" players where a player may represent a single packet (or a single session) in a context where there is a huge population of packets (or sessions). The result of such a decentralized decision making can be expected to satisfy the following properties which define the so called user (or Wardrop) equilibrium:

"*Under equilibrium conditions traffic arranges itself in congested networks such that all used routes between an OD pair (origin-destination pair) have equal and minimum costs while all unused routes have greater or equal costs*" [2].

Motivation.- A popular objective in some routing protocols in ad hoc networks is to assign routes for packets in such a way that each packet follows a minimal cost path (given the others' paths choices) [54]. This has the advantage of equalizing origin-destination delays of packets that belong to the same class, which allows one to minimize the amount of packets that come out of the sequence (this is desirable since in data transfers, out of order packets are misinterpreted to be lost which results not only in retransmissions but also in drop of system throughput).

Related work.- Both the framework of system-optimization as well as the minimum cost path have been studied extensively in the context of road traffic engineering. The use of a continuum network approach was already introduced on 1952 by Wardrop [2] and by Beckmann [28]. For more recent papers in this area, see *e.g.* [29, 30, 31, 32, 33] and references therein. We formulate it below and obtain some of its properties.

Congestion dependent cost.- We allow the local transmission cost c_1 for a horizontal transmission (in the direction of the axis x_1) to be different than the individual transmission cost c_2 for a vertical transmission (in the direction of the axis x_2). We add to the individual transmission cost c_1 the dependence on the traffic flow T_1 (in the direction of the axis x_1) and to the individual transmission cost c_2 the dependence on the traffic flow T_2 (in the direction of the axis x_2).

Let $V^k(\mathbf{x})$ be the minimum cost to go from a point $\mathbf{x}$ to B^k at equilibrium. Equation (3.7) still holds but this time with c_1 and c_2 that depends on T_1^k, T_2^k, and on the total flows T_1, T_2. Thus (3.8) becomes, for all $k \in \{1, \ldots, m\}$,

$$0 = \min_{i=1,2} \left(c_i(\mathbf{x}, T_i) + \frac{\partial V^k(\mathbf{x})}{\partial x_i} \right), \quad \forall \mathbf{x} \in B^k \, , V^k(\mathbf{x}) = 0 \, . \tag{3.9}$$

Notice that this method can be viewed as a generalization of the optimization method known as dynamic programming, in particular, last equation would be a generalization of Bellman equation also known as dynamic programming equation.

We note that if $T_i^k(\mathbf{x}) > 0$ then by the definition of the equilibrium, i attains the minimum at (3.9). Hence (3.9) implies the following relations for each traffic class k, and for $i = 1, 2$:

$$c_i(\mathbf{x}, T_i) + \frac{\partial V^k}{\partial x_i} = 0 \quad \text{if} \quad T_i^k > 0, \tag{3.10a}$$

$$c_i(\mathbf{x}, T_i) + \frac{\partial V^k}{\partial x_i} \geq 0 \quad \text{if} \quad T_i^k = 0. \tag{3.10b}$$

This is a set of coupled PDE's (Partial Differential Equations), actually difficult to analyze further.

Beckmann transformation

As Beckmann *et al.* did in [7] for discrete networks, we transform the minimum cost path problem into an equivalent system-minimization one. We shall restrict our analysis to the single class case. To that end, we note that equations (3.10a)-(3.10b) have exactly the same form as the Karush-Kuhn-Tucker conditions (3.1a)-(3.1b), except that $c_i(\mathbf{x}, T_i)$ in the former are replaced by $\partial g(\mathbf{x}, \mathbf{T})/\partial T_i(\mathbf{x})$ in the latter. We therefore introduce a *potential function* ψ defined by

$$\psi(\mathbf{x}, \mathbf{T}) = \sum_{i=1,2} \int_0^{T_i} c_i(\mathbf{x}, s) \mathrm{d}s$$

so that for both $i = 1, 2$:

$$c_i(\mathbf{x}, T_i) = \frac{\partial \psi(\mathbf{x}, \mathbf{T})}{\partial T_i}.$$

Then the user equilibrium flow is the one obtained from the system-optimization problem where we use $\psi(\mathbf{x}, \mathbf{T})$ as local cost. We conclude the following:

Theorem 3.5.1 *Let x^* be a solution to the following system-optimization problem.*

$$\min_{T(\cdot)} \int_{\mathcal{D}} \psi(\mathbf{x}, \mathbf{T})\, d\mathbf{x} \qquad \textit{subject to } \nabla \cdot \mathbf{T}(\mathbf{x}) = \rho(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathcal{D}.$$

Then it is the Wardrop equilibrium.

Remark 3.5.1 *In the special case where costs are given as a power of the flow as defined in Eq. (3.2), we observe that equations (3.10a)-(3.10b) coincide with equations (3.3a)-(3.3b) (up-to a multiplicative constant of the cost). We conclude that for such costs, the user equilibrium and the system-optimization solution coincide.*

Example 3.5.1 *The following example is an adaptation of the road traffic problem solved by Dafermos in [29] to our ad hoc setting. We therefore use the notation of [29] for the orientation, as we did in Section 3. Thus the direction from North-to-South will be our positive x_1 axis, and from West-to-East will be the positive x_2 axis. The framework we study is the user optimization problem with congestion dependent cost. For each point on the West and/or North boundary we consider the point to boundary problem. We thus seek a Wardrop equilibrium where each user can choose its destination among a given set. A flow configuration is a Wardrop equilibrium if under this configuration, each origin chooses a destination and a path to that destination that minimize that user's cost among all its possible choices.*

Consider the rectangular area R on the bounded domain $\mathcal{D}$ defined by the simple closed curve $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$ where

$$\begin{aligned} \Gamma_1 &= \{0 \le x_1 \le a, \quad x_2 = 0\}, \quad \Gamma_2 = \{x_1 = a, \quad 0 \le x_2 \le b\}, \\ \Gamma_3 &= \{0 \le x_1 \le a, \quad x_2 = b\}, \quad \Gamma_4 = \{x_1 = 0, \quad 0 \le x_2 \le b\}. \end{aligned}$$

Assume throughout that $\rho = 0$ for all $\mathbf{x}$ in the interior of $\mathcal{D}$, and that the costs of the routes are linear, i.e.,

$$c_1 = k_1 T_1 + h_1 \quad \textit{and} \quad c_2 = k_2 T_2 + h_2, \tag{3.11}$$

with $k_1 > 0$, $k_2 > 0$, h_1, and h_2 constant over $\mathcal{D}$. Linear costs can be viewed as a Taylor approximation of an arbitrary cost.

We are precisely in the framework of Section 3 and Section 3.5 with affine costs per packet. As a matter of fact, the potential function associated with these costs is

$$\psi(\mathbf{T}) = \sum_{i=1}^{2} \int_0^{T_i} (k_i s + h_i)\, ds = \sum_{i=1}^{2} \left(\frac{1}{2} k_i T_i + h_i \right) T_i.$$

Now, we want to handle a condensation of origins or destinations along the boundary. While this is feasible with the framework of section 3, it is rather technical. We rather use a more direct path below.

Notice that in the interior of $\mathcal{D}$, we have

$$\frac{\partial T_1}{\partial x_1} + \frac{\partial T_2}{\partial x_2} = 0.$$

Take any closed path γ surrounding a region ω. Then by Green formula,

$$\oint_\gamma T_1 d\mathcal{P}_2 - T_2 d\mathcal{P}_1 = \int_\omega \frac{\partial T_1}{\partial x_1} + \frac{\partial T_2}{\partial x_2} = 0$$

Therefore we can define

$$\phi(\mathbf{x}) := \int_{\mathbf{x^o}}^{\mathbf{x}} T_1 d\mathcal{P}_2 - T_2 d\mathcal{P}_1$$

the integral will not depend on the path between $\mathbf{x^o}$ and $\mathbf{x}$ and ϕ is thus well defined, and we have

$$\frac{\partial \phi(\mathbf{x})}{\partial x_2} = T_1(\mathbf{x}) \qquad \frac{\partial \phi(\mathbf{x})}{\partial x_1} = -T_2(\mathbf{x})\,. \tag{3.12}$$

We now make the assumption that there is sufficient demand and that the congestion cost is not too high so that at equilibrium the traffic T_1 and T_2 are strictly positive over all $\mathcal{D}$ [29]. It turns out that all paths to the destination are used. Thus, from Wardrop's principle, the cost $\int \mathbf{c}\, dx$ is equalized between any two paths. And therefore,

$$\frac{\partial c_1}{\partial x_2} = \frac{\partial c_2}{\partial x_1}.$$

Using the equations in (5.7b) then

$$k_1 \frac{\partial T_1}{\partial x_2} = k_2 \frac{\partial T_2}{\partial x_1},$$

and from equations in (5.7a) we have

$$k_1 \frac{\partial^2 \phi}{\partial x_2^2} + k_2 \frac{\partial^2 \phi}{\partial x_1^2} = 0.$$

Let $k_i = K_i^2$. Divide the above equation by $k_1 k_2$. One obtains

$$\frac{1}{K_1^2} \frac{\partial^2 \phi}{\partial x_1^2} + \frac{1}{K_2^2} \frac{\partial^2 \phi}{\partial x_2^2} = 0.$$

Following the classical way of analyzing the Laplace equation, (see [55]) we attempt a separation of variables according to

$$\phi(x_1, x_2) = F_1(K_1 x_1) F_2(K_2 x_2)\,.$$

We then get that

$$\frac{F_1''(K_1x_1)}{F_1(K_1x_1)} = -\frac{F_2''(K_2x_2)}{F_2(K_2x_2)} = s^2\,.$$

In that formula, since the first term is independent on x_2 and the second on x_1, then both must be constant. We call s^2 that constant, but we do not know its sign. Therefore, s may be imaginary or real. All solutions of this system for a given s are of the form

$$F_1(x) = A\cos(isx) + B\sin(isx)\,, \qquad F_2 = C\cos(sx) + D\sin(sx)\,.$$

As a matter of fact, ϕ may be the sum of an arbitrary number of such multiplicative decompositions with different s. We therefore arrive at the general formula

$$\phi(x_1, x_2) = \int [A(s)\cos(isK_1x_1) + B(s)\sin(isK_1x_1)][C(s)\cos(sK_2x_2) + D(s)\sin(sK_2x_2)]\,ds.$$

From this formula, we can write T_1 and T_2 as integrals also. The flow T at the boundaries should be orthogonal to the boundary, and have the local origin density for inward modulus (it is outward at a sink). It remains to expand these boundary conditions in Fourier integrals to identify the functions A, B, C, and D, which is tedious but straightforward (it is advisable to represent the integrals of the boundary densities as Fourier integrals, since then the boundary conditions themselves will be of the form $s\int R(s)\,ds$, closely matching the formulas we obtain for the T_i's).

Chapter 4

Omni-directional Antennas

Inspired by the work of Dafermos [29], who considered the routing problem over two possible directions (North-to-South and West-to-East), we have studied in the previous chapter the routing problem in massively dense ad hoc networks with directional antennas. In this chapter, we study the case where any general direction can be chosen at any location. Within this context, we analyze the system-optimization and the user-optimization problem. Afterward, we further study some important examples and give some comments and remarks about the results obtained in this chapter.

4.1 Introduction

An important approach for routing in wireless ad hoc networks has been to design traffic dependent adaptive protocols that send packets along paths that have smallest delays. This metrics goes back to an early paper by Gupta and Kumar [54] who show that by doing so, resequencing delays (that are undesirable in real time traffic and that are very harmful in data transfers using the TCP protocol) are minimized. A recent line of research has been to study such protocols in massively dense static wireless ad hoc networks.

Two types of objectives are aimed for the routing problem in the road traffic context. The first is to minimize the total cost of the system and the second is to find a routing configuration (called "traffic assignment") such that each transmission uses only paths with minimum costs. Configurations satisfying this property are known as "Wardrop Equilibrium", and they coincide with the solution concept used by Gupta and Kumar [54]. We study the two types of objectives in this paper in the context of massively dense static wireless ad hoc networks. For the first objective (which corresponds to cooperation between nodes) we use and strengthen results of Beckmann by using tools from optimization and control theory that were not available at the middle of the last century. We further study the Wardrop equilibrium and establish conditions under which it coincides with the system-optimization solution.

After describing the model in next section, we provide in Section 4.3 the mathematical foundations for the system-optimization problem. The mathematical foundations for describing and solving the user-optimization (*i.e.* the Wardrop equilibrium) are introduced in Section 4.4. This is followed by Section 4.5 with some important examples for the congestion cost. Finally, we end this chapter with a concluding section that summarizes our contributions in this context.

4.2 Routing in massively dense static ad hoc networks with omni-directional antennas

We recall that we are interested on the routing problem in a massively dense ad hoc network. We consider, within this context, a grid area network or domain $\mathcal{D}$ of the two-dimensional plane, with boundary $\mathcal{S}$, densely covered by potential relay nodes. In the previous chapter, messages could be transmitted on two possible directions (North-to-West or West-to-East) and that was justified by the use of directional antennas. In this chapter, messages can be transmitted on any direction at any location, but we restrict ourselves to the case where the sources and destinations are located within the boundary. As we will see in the examples of the following chapter, this is a reasonable assumption for many interesting cases. Under this assumption, messages have to be transmitted from a region of sources or origins of the information $\mathcal{O}$ to a disjoint region of receivers of the information $\mathcal{R}$ (in wireless sensor networks, it would correspond to data aggregation centers). Both of these regions are assumed to be located at disjoint portions of the boundary. The intensity $\sigma(x_1, x_2)$ of message generation on $\mathcal{O}$ is given, while the intensity $\rho(x_1, x_2)$ of signal destination on $\mathcal{R}$ is unknown. It is only assumed that these are consistent: the total transmission of messages emitted and received are equal. On the rest of the boundary (denoted by $\mathcal{F}$), no message should enter or leave $\mathcal{D}$, *i.e.* it is a forbidden-to-cross region. The congestion cost per packet transmitted (say in terms of delays, or energy use) at each location in $\mathcal{D}$ is a function $c(x_1, x_2, \varphi)$ of the location and of the intensity φ of the messages transmission through that point. We wish to investigate the system-optimal routing policy and its relationship with a Wardrop (or user-) kind of optimality.

Formal equations

Let $\mathcal{D}$ be an grid area network or domain of the plane $\mathbb{R}^2$ with smooth boundary $\mathcal{S}$. We assume that the domain $\mathcal{D}$ is at every point of the boundary $\mathcal{S}$ on a single side of $\mathcal{S}$, so that an exterior normal to $\mathcal{D}$, say $\mathbf{n}(\mathbf{x})$ is well defined and smooth on $\mathcal{S}$. This last assumption is made to avoid "strange" boundaries that are unrealistic in practice.

We model the transmission of messages as a vector field $\mathbf{T} : \mathcal{D} \to \mathbb{R}^2$, and we let $\varphi(\mathbf{x}) = \|\mathbf{T}(\mathbf{x})\|$ be its intensity. The transmission of messages through $\mathcal{O}$ is given as a continuously differentiable function $\sigma(\cdot) : \mathcal{O} \to \mathbb{R}_+$. The conservation assumption now reads

$$\int_{\mathcal{R}} \rho(\mathbf{x})\, ds = \int_{\mathcal{O}} \sigma(\mathbf{x})\, ds\,. \tag{4.1}$$

Let $\mathcal{Q} = \mathcal{O} \cup \mathcal{F}$ and extend the function σ to the whole of $\mathcal{Q}$ by $\sigma(\mathbf{x}) = 0$ on $\mathcal{F}$. We model the conditions on the boundary as

$$\forall \mathbf{x} \in \mathcal{Q}\,, \quad \langle \mathbf{n}(\mathbf{x}), \mathbf{T}(\mathbf{x}) \rangle = -\sigma(\mathbf{x}). \tag{4.2}$$

There is neither source nor destination of messages inside $\mathcal{D}$, which we model as the constraint (see Section 2.2, Eq. (2.7)):

$$\forall \mathbf{x} \in \mathcal{D}\,, \quad \nabla \cdot \mathbf{T}(\mathbf{x}) = 0\,. \tag{4.3}$$

It follows that

$$\int_{\mathcal{S}} \langle \mathbf{n}(\mathbf{x}), \mathbf{T}(\mathbf{x}) \rangle \, ds = 0\,,$$

which suffices to insure the consistency condition (4.1).

The congestion cost per packet c is supposed to be a strictly positive continuously differentiable function $c(\mathbf{x}, \varphi) : \mathcal{D} \times \mathbb{R}_+ \to \mathbb{R}_+$, increasing and convex in φ for each $\mathbf{x}$. The total cost of congestion will be taken as

$$G(\mathbf{T}(\cdot)) = \int_{\mathcal{D}} c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)) \cdot \|\mathbf{T}(\mathbf{x}))\| \; d\mathbf{x}. \tag{4.4}$$

The path followed by a packet is specified by its direction of travel $e_\theta = (\cos\theta,\, \sin\theta)$ along its path, according to $\dot{\mathbf{x}} = e_\theta$. The cost incurred by one packet traveling from $\mathbf{x}_0 \in \mathcal{O}$ at time t_0 to $\mathbf{x}_1 \in \mathcal{R}$ reached at time t_1 is

$$J(e_\theta(\cdot)) = \int_{\mathbf{x}_0}^{\mathbf{x}_1} c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)\sqrt{dx^2 + dy^2} = \int_{t_0}^{t_1} c(\mathbf{x}(t), \|\mathbf{T}(\mathbf{x}(t))\|) \; dt\,. \tag{4.5}$$

Notice that this "time" t may be a fictitious time, related to physical time, say τ, by $d\tau = c\; dt$ for instance. Then c is the inverse of a speed of travel, a delay due to congestion, and J is the time taken by the message to go from source to destination.

Regularity and function spaces

We shall seek $\mathbf{T}(\cdot)$ in a space which we denote V. We next discuss the choice of the functional spaces so that the problem is a well-posed problem. A non-mathematical oriented reader may skip the description of the function spaces we introduce.

The mathematical term well-posed problem stems from a definition given by Hadamard [56]. He believed that mathematical models of physical phenomena should have the properties that

- a solution exists,
- the solution is unique,
- the solution depends continuously on the data, in some reasonable topology.

Notice that in order to obtain a well-posed problem, we may choose $V = (H^1(\mathcal{D}))^2$, but this will require $\sigma(\cdot)$ to be slightly more regular than necessary, namely, $H^{1/2}(\mathcal{S})$. To keep with the classical hypothesis in fluid dynamics, we may choose $V = (H_\nabla(\mathcal{D}))^2$, the space of L^2 functions whose divergence is in L^2. Then we choose $\sigma(\cdot)$ in $L^2(\mathcal{S})$.

The above Sobolev spaces have been introduced by the modern theory of partial differential equations (PDEs) [57]. An extensive theory of PDEs and their numerical approximations is now available in these Sobolev spaces. This choice of spaces allows one to have complete spaces for the considered functions and for their derivatives, along with a scalar product for square-integrable functions. The completeness of the space is needed to have existence of the solution. The scalar product allows to have duality. The completeness together with the duality allows Karush-Kuhn-Tucker Theorem to hold, which we make use of in this work.

Let V_0 be the closure in V of the set of C^∞ functions with compact support in $\mathcal{D}$. Let $V_\mathcal{R}$ and $V_\mathcal{Q}$ be the closures in V of the set of $\mathcal{C}^\infty$ functions that are null in a neighborhood of $\mathcal{R}$ and $\mathcal{Q}$ respectively. They are vector spaces, super-sets of V_0. Let $\mathbf{T}(\cdot) : \mathcal{D} \to \mathbb{R}^2$ be a vector field in V satisfying the constraint (4.2) (for instance a smooth extension of $\sigma(\mathbf{x})n(\mathbf{x})$). Let $\mathcal{V}$ be the affine space $\tilde{f} + V_\mathcal{Q}$. We shall also need the space $H^1_\mathcal{R}$ of functions of $H^1(\mathcal{D})$ whose trace on $\mathcal{R}$ is zero. Finally, we let $\mathcal{D}_0 = \{\mathbf{x} \mid f^\star(\mathbf{x}) = 0\}$, or more precisely, since $f^\star$ is not necessarily continuous, the largest open subset of $\mathcal{D}$ over which $\int_{\mathcal{D}_0} \|f^\star(\mathbf{x})\|^2 \, d\mathbf{x} = 0$.

The case of elastic traffic

Let's assume that we do not have to transmit the whole demand $\sigma(x)$ to the destination. We shall send less if there is congestion. The standard way to model that is first to define a utility $u(s)$ for having s units of information transmitted; we take $s(x) \leq \sigma(x)$. The new objective is to minimize the sum of $C(f) - U(s)$ where $U(s)$ is the integral of $u(s(x))$ over x.

One way to solve the problem is to define a new destination S. Then add an alternative route from each source to S; the cost to transmit f units from a source x to S is $-u(\sigma(x) - f)$. Thus instead of directly adding utilities to the optimization problem, they appear through costs of new routes that are added. The elastic routing problem is thus transformed into an equivalent routing problem with fixed demand. This transformation is standard, see [58, 59], and we shall not pursue it here.

4.3 System-optimization

4.3.1 The differentiable case

We seek here the vector field $f^\star \in (L^2(\mathcal{D}))^2$ satisfying the constraints (4.2) and (4.3) and minimizing $G(f)$. Let $C(\mathbf{x}, \varphi) = c(\mathbf{x}, \varphi)\varphi$. It is convex in φ and coercive (*i.e.* goes to infinity with φ). As a consequence, $\mathbf{T}(\cdot) \mapsto G(\mathbf{T}(\cdot))$ is continuous, convex and coercive. Moreover, the constraints are linear. Therefore an optimum exists, and we may apply the Karush-Kuhn-Tucker Theorem (see Section 3.2).

We dualize only the constraint (4.3) and look for $f^\star$ in $\mathcal{V}$. Let therefore $p(\cdot) \in L^2(\mathcal{D})$ be the dual variable, we let

$$\mathcal{L}(f,p) = \int_{\mathcal{D}} \Big(C(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|) + p(\mathbf{x}) \nabla \cdot \mathbf{T}(\mathbf{x}) \Big) \, d\mathbf{x} \, .$$

Using Green's formula, we may also write

$$\mathcal{L}(f,p) = \int_{\mathcal{D}} \Big(C(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|) - \langle \nabla p(\mathbf{x}), \mathbf{T}(\mathbf{x}) \rangle \Big) \, d\mathbf{x} + \int_{\mathcal{S}} p(\mathbf{x}) \langle n(\mathbf{x}), \mathbf{T}(\mathbf{x}) \rangle \, ds \, .$$

The optimal vector field $f^\star$ should minimize $\mathcal{L}$ over $\mathcal{V}$, for some p. Therefore, 0 must belong to the sub-differential with respect to f of the restriction of $\mathcal{L}$ to $\mathcal{V}$.

Wherever $f^\star \neq 0$, $\mathcal{L}$ is actually differentiable, so that the sub-differential contains only the derivative. Actually, we only need the restriction of the derivative to $V_{\mathcal{Q}}$.

$$\mathrm{D}\mathcal{L} \cdot g = \int_{\mathcal{D}} \Big(\mathrm{D}_2 C(\mathbf{x}, \|f^\star(\mathbf{x})\|) \frac{\langle f^\star(\mathbf{x}), g(\mathbf{x}) \rangle}{\|f^\star(\mathbf{x})\|} - \langle \nabla p(\mathbf{x}), g(\mathbf{x}) \rangle \Big) \, d\mathbf{x} + \int_{\mathcal{R}} p(\mathbf{x}) \langle n(\mathbf{x}), g(\mathbf{x}) \rangle \, ds \, ,$$

which should be zero for every $g \in V_{\mathcal{Q}}$. Pick first g in V_0. The last integral vanishes. It follows that necessarily

$$\forall \mathbf{x} : f^\star(\mathbf{x}) \neq 0 \, , \quad \mathrm{D}_2 C(\mathbf{x}, \|f^\star(\mathbf{x})\|) \frac{f^\star(\mathbf{x})}{\|f^\star(\mathbf{x})\|} = \nabla p(\mathbf{x}) \, . \tag{4.6}$$

It follows from this equation that $p(\cdot) \in H^1(\mathcal{D})$, and also that the first integral in the right-hand side must be zero for every g in $V_{\mathcal{Q}}$. Picking now $g \in V_{\mathcal{Q}}$, it follows that

$$p(\cdot) \in H^1_{\mathcal{R}}. \tag{4.7}$$

Wherever $\|f^\star(\mathbf{x})\| = 0$, a discussion arises. If $D_2 C(\mathbf{x}, \varphi)/\varphi$ remains bounded as $\varphi \to 0$, there is nothing to add to equations (4.6) and (4.7) above. (We shall see the typical example $C(\mathbf{x}, \varphi) = (1/2) c(\mathbf{x}) \varphi^2$ below.) Otherwise the situation is more complicated.

4.3.2 Lack of differentiability

We investigate now the case where $\mathrm{D}_2 C(\mathbf{x}, \varphi)/\varphi \to \infty$ as $\varphi \to 0$. This typically arises, *e.g.* if $\mathrm{D}_2 C(\mathbf{x}, 0) \neq 0$. We shall see the typical example $C(\mathbf{x}, \varphi) = c(\mathbf{x}) \varphi$ below. Then, $f \mapsto C(\mathbf{x}, \|f\|)$ is not differentiable (with respect to f) at 0. Its sub-differential is the set

$$\partial_f C(\mathbf{x}, 0) = \{ q \in \mathbb{R}^2 \mid \forall g \in \mathbb{R}^2 \, , \; C(\mathbf{x}, \|g\|) - C(\mathbf{x}, 0) \geq \langle q, g \rangle \} \, .$$

Since C is assumed to be differentiable and convex in its second argument, this is equivalent to

$$\partial_f C(\mathbf{x}, 0) = \{ q \mid \forall g \in \mathbb{R}^2 \, , \; \mathrm{D}_2 C(\mathbf{x}, 0) \|g\| \geq \langle q, g \rangle \} \, ,$$

which in turn is equivalent to $\|q\| \leq |\mathrm{D}_2C(\mathbf{x},0)|$. Now, since C is assumed to be increasing in φ, $\mathrm{D}_2C \geq 0$. Placing this back into the sub-differential of $\mathcal{L}$, we get, for $\mathbf{x} \in \mathcal{D}_0$,

$$\exists q(\mathbf{x}) \text{ such that } \|q(\mathbf{x})\| \leq \mathrm{D}_2C(\mathbf{x},0) \text{ and } \forall g \in V_{\mathcal{Q}}\,, \int_{\mathcal{D}_0} (q(\mathbf{x}) - \nabla p(\mathbf{x}))g(\mathbf{x})\, d\mathbf{x} = 0\,.$$

Combining both cases, we conclude that, for a function $f^\star(\cdot) \in V$ with null set $\mathcal{D}_0$ to be optimal, there must exist a $p(\cdot) \in H^1_{\mathcal{R}}$ such that

$$\begin{aligned} &\forall \mathbf{x} \in \mathcal{D}\,, && \|\nabla p(\mathbf{x})\| \leq \mathrm{D}_2C(\mathbf{x},0)\,, \\ &\forall \mathbf{x} \in \mathcal{D} - \mathcal{D}_0\,, && \nabla p(\mathbf{x}) = \mathrm{D}_2C(\mathbf{x}, \|\mathbf{T}(\mathbf{x})^\star\|)\tfrac{1}{\|f^\star(\mathbf{x})\|} f^\star(\mathbf{x})\,. \end{aligned} \tag{4.8}$$

We may notice that the first condition above also yields

$$\forall \mathbf{x} : f^\star(\mathbf{x}) \neq 0\,, \quad \|\nabla p(\mathbf{x})\| = D_2C(\mathbf{x}, \|f^\star(\mathbf{x})\|).$$

Overall, the problem of determining the optimum $f^\star$ is equivalent (if that system has a single solution) to determining simultaneously $f^\star$ and p satisfying (4.2),(4.3) and (4.8).

This system certainly has at least one solution, since our problem is convex, coercive with affine constraints, and thus has a minimum. Uniqueness, on the other hand, is by no means simple. It may be noticed that one might look for the two scalar functions φ and p, satisfying

$$\begin{aligned} &\forall \mathbf{x} : \varphi(\mathbf{x}) \neq 0\,, && \|\nabla p(\mathbf{x})\| = \mathrm{D}_2C(\mathbf{x}, \varphi(\mathbf{x}))\,, \\ &\forall \mathbf{x} : \varphi(\mathbf{x}) = 0\,, && \|\nabla p(\mathbf{x})\| \leq \mathrm{D}_2C(\mathbf{x}, 0)\,, \\ &\forall \mathbf{x} \in \mathcal{R}\,, && p(\mathbf{x}) = 0\,, \end{aligned}$$

and impose furthermore the constraints (4.2) and (4.3) on

$$f^\star(x) = \frac{\varphi(\mathbf{x})}{\mathrm{D}_2C(\mathbf{x}, \varphi(\mathbf{x}))} \nabla p(\mathbf{x})\,.$$

We shall investigate a typical case hereafter.

4.4 User-optimization (Wardrop equilibrium)

Assume the message flow obeys the above necessary conditions. We want to investigate whether it is optimal for a single message to follow the route prescribed by $f^\star$, *i.e.*, an integral line of that field, assuming that its lone deviation from that scheme would have no effect on the overall congestion of the network.

We investigate the optimization of the criterion (5.4) via its Hamilton-Jacobi-Bellman equation. Let $V(\mathbf{x})$ be the return function, it must be a viscosity solution of

$$\begin{aligned} &\forall \mathbf{x} \in \mathcal{D}\,, && \min_\theta \langle e_\theta, \nabla V(\mathbf{x})\rangle + c(\mathbf{x}, \|f^\star(\mathbf{x})\|) = 0\,, \\ &\forall \mathbf{x} \in \mathcal{R}\,, && V(\mathbf{x}) = 0\,. \end{aligned}$$

hence

$$\begin{aligned} &\forall \mathbf{x} \in \mathcal{D}, \quad -\|\nabla V(\mathbf{x})\| + c(\mathbf{x}, \|f^\star(\mathbf{x})\|) = 0\,, \\ &\forall \mathbf{x} \in \mathcal{R}, \quad V(\mathbf{x}) = 0\,. \end{aligned} \tag{4.9}$$

And the optimal direction of travel is opposite to $\nabla V(\mathbf{x})$, *i.e.*, $e_\theta = -\nabla V(\mathbf{x})/\|\nabla V(\mathbf{x})\|$.

Clearly, this is the same system of equations as previously, upon replacing $p(\mathbf{x})$ by $-V(\mathbf{x})$, and $\mathrm{D}_2 C(\mathbf{x}, \varphi)$ by $c(\mathbf{x}, \varphi)$. We thus conclude that the Wardrop equilibrium can be obtained by solving the globally optimal problem in which the cost density is replaced by $\int_0^\varphi c(\mathbf{x}, s) ds$. This is the continuous version of the potential function approach of Beckmann *et al.* [7]. This transformation has been frequently used in the road traffic context but only for one particular cost structure [37, 38, 33, 39] the equivalence was shown to hold in [37, 38].

4.5 Important examples

4.5.1 Monomial cost

In the case where $c(\mathbf{x}, \varphi) = c(\mathbf{x})\varphi^\alpha$, then $C(\mathbf{x}, \varphi) = \alpha c(\mathbf{x}, \varphi)$, and therefore the two systems of equations coincide (up to a constant), or more precisely, they coincide in the domain $\{\mathbf{x} \mid f^\star(\mathbf{x}) \neq 0\}$. We shall show that for a given $\varphi(\cdot)$, p is uniquely defined. We therefore have the following property:

Proposition 4.5.1 *For a monomial cost, any global equilibrium where $\mathcal{D}_0 = \emptyset$ is a Wardrop equilibrium.*

4.5.2 Linear congestion cost

We investigate here the simple typical case, where the cost of congestion is linear:

$$c(\mathbf{x}, \varphi) = \frac{1}{2} c(\mathbf{x})\varphi, \text{ so that } C(\mathbf{x}, \varphi) = \frac{1}{2} c(\mathbf{x})\varphi^2\,.$$

Then, $\mathcal{L}$ is differentiable everywhere, and the necessary condition of optimality is just that there should exist $p : \mathcal{D} \to \mathbb{R}^2$ such that $\nabla p(\mathbf{x}) = c(\mathbf{x}) f^\star(\mathbf{x})$. Placing this into (4.2) and (4.3), we see that we end up with a simple elliptic equation with mixed Dirichlet - (non-homogeneous) Neumann boundary conditions:

$$\left.\begin{aligned} &\forall \mathbf{x} \in \mathcal{D}, \quad \nabla\left(\tfrac{1}{c(\mathbf{x})}\nabla p(\mathbf{x})\right) = 0\,, \\ &\forall \mathbf{x} \in \mathcal{Q}, \quad \frac{\partial p}{\partial n}(\mathbf{x}) = c(\mathbf{x})\sigma(\mathbf{x})\,, \\ &\forall \mathbf{x} \in \mathcal{R}, \quad p(\mathbf{x}) = 0\,, \end{aligned}\right\} \tag{4.10}$$

for which existence and uniqueness of the solution follows.

A more or less explicit solution can then be given in terms of the Green function $\mathcal{G}(\mathbf{x}, \mathcal{P})$ of the domain

$$f^\star(\mathbf{x}) = \int_{\mathcal{Q}} \frac{1}{c(\mathbf{x})} \nabla_1 \mathcal{G}(\mathbf{x}, \mathcal{P})\sigma(\mathcal{P})\, ds(\mathcal{P})\,.$$

If the Green function is not available, according to a classical approach, we may derive a finite element method from the variational form: Find $p \in H^1_{\mathcal{R}}$ such that, for any $q \in H^1_{\mathcal{R}}$,

$$\int_{\mathcal{D}} \frac{1}{c(\mathbf{x})} \langle \nabla p(\mathbf{x}), \nabla q(\mathbf{x}) \rangle \, dx - \int_{\mathcal{Q}} \sigma(\mathbf{x}) q(\mathbf{x}) \, ds = 0 \, .$$

This can be read as $\mathrm{D}K(p) = 0$ where $K : H^1_{\mathcal{R}} \to \mathbb{R}$ is given by

$$K(p) = \frac{1}{2} \int_{\mathcal{D}} \frac{1}{c(\mathbf{x})} \|\nabla p(\mathbf{x})\|^2 - \int_{\mathcal{Q}} \sigma(\mathbf{x}) p(\mathbf{x}) \, ds \, .$$

Thanks to Poincaré's inequality, it is convex and coercive. We therefore obtain:

Proposition 4.5.2 *Equations (4.10) have a unique solution* $p \in H^1_{\mathcal{R}}$.

4.5.3 Uncongested network

An algorithm

We consider now a situation where the network operates far from congestion. The "cost" $c(\mathbf{x})$ may be regarded as a delay, then the cost of any trajectory is just the time it takes, or an energy expenditure. In any case, it is related to the state of the infrastructure, not to its load. Then, c is independent of $\|\mathbf{T}(\mathbf{x})\|$, and we get $C(\mathbf{x}, \varphi) = c(\mathbf{x})\varphi$. Then, (4.8) simplifies into

$$\begin{aligned} &\forall \mathbf{x} \in \mathcal{D} \, , \quad \|\nabla p(\mathbf{x})\| \le c(\mathbf{x}) \, , \\ &\forall \mathbf{x} : f^\star(\mathbf{x}) \neq 0 \, , \quad \nabla p(\mathbf{x}) = c(\mathbf{x}) \frac{f^\star(\mathbf{x})}{\|f^\star(\mathbf{x})\|} \, . \end{aligned}$$

Let

$$\varphi(\mathbf{x}) = \|f^\star(\mathbf{x})\| \, , \quad \psi(\mathbf{x}) = \frac{\varphi(\mathbf{x})}{c(\mathbf{x})} \, .$$

The above system yields

$$\forall \mathbf{x} \in \mathcal{D} \, , \quad \psi(\mathbf{x}) \ge 0 \, , \quad \|\nabla p(\mathbf{x})\| \le c(\mathbf{x}) \, , \quad \psi(\mathbf{x})[\|\nabla p(\mathbf{x})\| - c(\mathbf{x})] = 0 \, , \tag{4.11}$$

and also $f^\star(\mathbf{x}) = \psi(\mathbf{x}) \nabla p(\mathbf{x})$, which placed in (4.3) and (4.2) yields

$$\begin{aligned} &\forall \mathbf{x} \in \mathcal{D} \, , \quad \psi(\mathbf{x}) \Delta p(\mathbf{x}) + \langle \nabla \psi(\mathbf{x}), \nabla p(\mathbf{x}) \rangle = 0 \, , \\ &\forall \mathbf{x} \in \mathcal{S} \, , \quad \psi(\mathbf{x}) \langle n(\mathbf{x}), \nabla p(\mathbf{x}) \rangle = \sigma(\mathbf{x}) \, . \end{aligned} \tag{4.12}$$

We do not have a satisfactory theory of this equation. Even if, as we noticed, the existence is guaranteed, we do not know whether that solution is unique. It should be noticed that the uniqueness proof given for a very similar equation in [28] does not carry over here, because it relies critically on the strict convexity of the cost in $\|f\|$.

As an attempt, we provide here an iterative algorithm which, if it converges, converges toward a solution of the system. It provides us with a uniqueness result under a strong hypothesis. We suspect that a more general result is true, and also that the algorithm converges even without that hypothesis.

We seek ψ in $H^1(\mathcal{D})$, and p in $H^1_{\mathcal{R}}$. We may reformulate the system (4.12) as $\forall q \in V_{\mathcal{R}}$,

$$\int_{\mathcal{D}} [\psi(\mathbf{x})\Delta p(\mathbf{x}) + \langle \nabla\psi(\mathbf{x}), \nabla p(\mathbf{x})\rangle] q(\mathbf{x}) \; d\mathbf{x} - \int_{\mathcal{Q}} [\psi(\mathbf{x})\langle n(\mathbf{x}), \nabla p(\mathbf{x})\rangle - \sigma(\mathbf{x})] q(\mathbf{x}) \, ds = 0\,.$$

Using Green's formula for $q \in H^1(\mathcal{D})$:

$$\int_{\mathcal{D}} [\psi(\mathbf{x})\Delta p(\mathbf{x}) + \langle \nabla\psi(\mathbf{x}), \nabla p(\mathbf{x})\rangle] q(\mathbf{x}) \; d\mathbf{x} = \\ - \int_{\mathcal{D}} \psi(\mathbf{x})\langle \nabla p(\mathbf{x}), \nabla q(\mathbf{x})\rangle \; d\mathbf{x} + \int_{\mathcal{S}} \psi(\mathbf{x})\langle n(\mathbf{x}), \nabla p(\mathbf{x})\rangle q(\mathbf{x}) \; ds\,,$$

system (4.12) can therefore be stated as:

$$\forall q \in V_{\mathcal{R}}\,, \quad \int_{\mathcal{D}} \psi(\mathbf{x})\langle \nabla p(\mathbf{x}), \nabla q(\mathbf{x})\rangle \; d\mathbf{x} - \int_{\mathcal{Q}} \sigma(\mathbf{x}) q(\mathbf{x}) \; ds = 0\,. \tag{4.13}$$

This equality may also be interpreted as $\mathrm{D}_1 J(p, \psi) q = 0$ where $J : V_{\mathcal{R}} \to \mathbb{R}$ is defined by

$$J(p, \psi) = \frac{1}{2} \int_{\mathcal{D}} \psi(\mathbf{x}) \|\nabla p(\mathbf{x})\|^2 \; d\mathbf{x} - \int_{\mathcal{Q}} \sigma(\mathbf{x}) p(\mathbf{x}) \; ds\,.$$

Poincaré's inequality states that there exists $C > 0$ such that,

$$\forall p \in V_{\mathcal{R}}, \|p\|^2 \leq C \|\nabla p\|^2\,. \tag{4.14}$$

Thus the functional J above is coercive and has a single minimum.

One may guess the following algorithm: fix $\psi^0(\mathbf{x})$ (say $= 1$). Given ψ^n, minimize J with respect to p, say solving the finite element equations corresponding to (4.13). Call p^n the solution, and do

$$\psi^{n+1}(\mathbf{x}) = \max\{0, \psi^n(\mathbf{x}) + \theta(\|\nabla p^n(\mathbf{x})\|^2 - c(\mathbf{x})^2)\} \tag{4.15}$$

for some positive θ. We shall prove the following theorem :

Proposition 4.5.3 *If there exists a solution of equations (4.11)–(4.12) such that $\|f^\star\|$ is essentially bounded away from 0 in $\mathcal{D}$, it is unique and for θ small enough algorithm (4.15) converges toward that solution.*

Analysis of the algorithm

Let $\psi^\star$, $p^\star$ be a solution of our system of equations. Notice first that indeed, for any $\theta > 0$,

$$\forall \mathbf{x} \in \mathcal{D}\,, \quad \psi^\star(\mathbf{x}) = \max\{0, \psi^\star(\mathbf{x}) + \theta(\|\nabla p^\star(\mathbf{x})\|^2 - c(\mathbf{x})^2)\} \tag{4.16}$$

And any limit of the above algorithm has to satisfy this equation, which says that $\|\nabla p(\mathbf{x})\| = c(\mathbf{x})$ for every $\mathbf{x}$ where $\psi(\mathbf{x}) \neq 0$. Together with the condition that p minimizes J for ψ, this is exactly the conditions (4.11) and (4.12).

Subtract (4.16) from (4.15). It results that

$$|\psi^{n+1}(\mathbf{x}) - \psi^\star(\mathbf{x})| \leq |\psi^n(\mathbf{x}) - \psi^\star(\mathbf{x}) + \theta(\|\nabla p^n(\mathbf{x})\|^2 - \|\nabla p^\star(\mathbf{x})\|^2)|\,.$$

Take the square, and integrate over $\mathcal{D}$:

$$\begin{aligned}&\int_{\mathcal{D}}|\psi^{n+1}(\mathbf{x}) - \psi^\star(\mathbf{x})|^2\ d\mathbf{x} \leq \int_{\mathcal{D}}|\psi^n(\mathbf{x}) - \psi^\star(\mathbf{x})|^2\ d\mathbf{x}\\ &+2\theta\int_{\mathcal{D}}(\psi^n(\mathbf{x}) - \psi^\star(\mathbf{x}))(\|\nabla p^n(\mathbf{x})\|^2 - \|\nabla p^\star(\mathbf{x}))\|^2)\ d\mathbf{x}\\ &+\theta^2\int_{\mathcal{D}}(\|\nabla p^n(\mathbf{x})\|^2 - \|\nabla p^\star(\mathbf{x})\|^2)^2\ d\mathbf{x}\,.\end{aligned} \tag{4.17}$$

Using Cauchy-Schwarz inequality, the last term is bounded from above by

$$\int_{\mathcal{D}}(\|\nabla p^n(\mathbf{x})\|^2 - \|\nabla p^\star(\mathbf{x})\|^2)^2\ d\mathbf{x} \leq$$

$$\int_{\mathcal{D}}\|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2\ d\mathbf{x}\int_{\mathcal{D}}\|\nabla(p^n(\mathbf{x}) + p^\star(\mathbf{x}))\|^2\ d\mathbf{x}\,.$$

Hence, assuming $\int_{\mathcal{D}}\|\nabla p^n(\mathbf{x})\|^2\ d\mathbf{x}$ remains bounded, there exists $a > 0$ such that

$$\int_{\mathcal{D}}(\|\nabla p^n(\mathbf{x})\|^2 - \|\nabla p^\star(\mathbf{x})\|^2)^2\ d\mathbf{x} \leq a\int_{\mathcal{D}}\|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2\ d\mathbf{x}\,. \tag{4.18}$$

Concerning the second term of the right-hand side of (4.17), write

$$\|\nabla p^\star\|^2 = \|\nabla p^n + \nabla(p^\star - p^n)\|^2 = \|\nabla p^n\|^2 + 2\langle\nabla p^n, \nabla(p^\star - p^n)\rangle + \|\nabla(p^\star - p^n)\|^2\,.$$

Thus (using short notations for convenience)

$$\begin{aligned}&\frac{1}{2}\int_{\mathcal{D}}\psi^n\|\nabla p^\star\|^2\ d\mathbf{x} - \int_{\mathcal{Q}}\sigma p^\star\ ds\\ &= \frac{1}{2}\int_{\mathcal{D}}\psi^n\|\nabla p^n\|^2\ d\mathbf{x} - \int_{\mathcal{Q}}\sigma p^n\ ds + \frac{1}{2}\int_{\mathcal{D}}\psi^n\|\nabla(p^\star - p^n)\|^2\ d\mathbf{x}\\ &\qquad + \int_{\mathcal{D}}\psi^n\langle\nabla p^n, \nabla(p^\star - p^n)\rangle\ d\mathbf{x} - \int_{\mathcal{Q}}\sigma(p^\star - p^n)\ ds\,.\end{aligned}$$

By the definition of p^n as solving equation (4.13), the second line above is zero, leaving the first line alone. In a symmetric fashion, we also get

$$\frac{1}{2}\int_{\mathcal{D}}\psi^\star\|\nabla p^n\|^2 - \int_{\mathcal{Q}}\sigma p^n = \frac{1}{2}\int_{\mathcal{D}}\psi^\star\|\nabla p^\star\|^2 - \int_{\mathcal{Q}}\sigma p^\star + \frac{1}{2}\int_{\mathcal{D}}\psi^\star\|\nabla(p^n - p^\star)\|^2\,.$$

Summing the last two equalities (and multiplying by 2), we obtain

$$\int_{\mathcal{D}}(\psi^n - \psi^\star)(\|\nabla p^n\|^2 - \|\nabla p^\star\|^2) = -\int_{\mathcal{D}}(\psi^n + \psi^\star)\|\nabla(p^n - p^\star)\|^2\,.$$

Placing this and (4.18) in (4.17), we may summarize the above calculations as

$$
\begin{aligned}
&\int_{\mathcal{D}} |\psi^{n+1}(\mathbf{x}) - \psi^\star(\mathbf{x})|^2 \, d\mathbf{x} \le \int_{\mathcal{D}} |\psi^{n}(\mathbf{x}) - \psi^\star(\mathbf{x})|^2 \, d\mathbf{x} \\
&-2\theta \int_{\mathcal{D}} (\psi^n(\mathbf{x}) + \psi^\star(\mathbf{x})) \|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2 \, d\mathbf{x} \\
&+a\theta^2 \int_{\mathcal{D}} \|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2 \, d\mathbf{x} \,.
\end{aligned}
\tag{4.19}
$$

Assume that, for almost all $\mathbf{x} \in \mathcal{D}$, $\psi^\star(\mathbf{x}) \ge b > 0$. It follows that

$$
\int_{\mathcal{D}} (\psi^n(\mathbf{x}) + \psi^\star(\mathbf{x})) \|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2 \, d\mathbf{x} \ge b \int_{\mathcal{D}} \|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2 \, d\mathbf{x} \,,
$$

and therefore that for any $\theta \le b/a$,

$$
\int_{\mathcal{D}} |\psi^{n+1}(\mathbf{x}) - \psi^\star(\mathbf{x})|^2 \, d\mathbf{x} \le \int_{\mathcal{D}} |\psi^{n}(\mathbf{x}) - \psi^\star(\mathbf{x})|^2 \, d\mathbf{x} - b\theta \int_{\mathcal{D}} \|\nabla(p^n(\mathbf{x}) - p^\star(\mathbf{x}))\|^2 \, d\mathbf{x} \,.
$$

Summing these inequalities, it follows that the series of the L^2 norms $\|\nabla p^n \nabla p^\star\|^2$ converges, and according to Poincarés inequality again, $p^n \to p^\star$ in $H^1(\mathcal{D})$. The field of optimal directions converges as well, and assuming it is regular enough for the integral curves to be unique, the optimal field converges as well.

The algorithm is independent from the choice of $p^\star$ and $\psi^\star$ who are therefore uniquely defined.

4.6 Comments

We present a brief comparison of our treatment with the work made by Martin Beckman [28], called hereafter **M.B.**. In **M.B.**, one introduces both the density $u(\mathbf{x})$ of commodity to be moved, and the speed $v(\mathbf{x})$ of this motion, which is a data. And the cost of transportation is assumed to be a function of u alone. The decision variable in **M.B.** is the vector field φ of transportation where the direction of φ is that of the transportation, and $\|\varphi\|$ its density u. Hence **M.B.**'s $v\varphi$ is our f. And his equation (11) (see [28]) is our equation (4.6).

In **M.B.** there is an area source or destination of matter to be transported. It is not needed in our context, but technically, it would be trivially done just adding a non-zero right hand side to equation (4.3) and its various forms, the first equation of (4.10) and of (4.12).

Now, since the early 50's, the theory of partial differential equations (PDEs) has been considerably developed, using the tools of Sobolev spaces and the variational theory of J-L. Lions, P. Lax, and others. Thus our derivation is not formal any more, and we are able to give existence and uniqueness theorems impossible to derive in 1952 when his work was published. Notice that our example with no congestion, where our uniqueness theorem is not very satisfactory, does not satisfy the hypotheses of the uniqueness theorem of **M.B.**, because that paper requires that the cost function be strictly convex.

Finally, we solve for the concept of Wardrop equilibrium, and we are therefore able to compare the global optimum to the Wardrop equilibrium, which was not available to Beckmann in 1952.

By casting the routing problem in massively dense wireless ad hoc networks in the context of the road traffic framework of Beckmann, we are able to formulate and solve various optimization problems and study various cost functions, which was not the case with the physics-inspired paradigms that had been used before to study massively dense ad hoc networks.

Chapter 5

Numerical Analysis

The road traffic community and the partial differential equations community have developed numerical approaches to solve the continuous approximation model through the discretization of the region under study. Although it may seem that one is back to the starting point with yet another discrete problem to solve, the new discrete problem is simpler, and the resolution is independent of the number of nodes in the original system. In this chapter, we analyze the numerical resolution to the routing problem in massively dense static wireless ad hoc networks via Finite Element Method (FEM). We focus in some examples where we are able to solve the system- and the user-optimization problem.

5.1 Model specification

We recall some of the notations used and results obtained in the previous chapter:

A domain $\mathcal{D}$ of the two-dimensional plane is densely covered by relay nodes. Messages have to flow from a region of origins of the information $\mathcal{O}$ to a region of receivers of this information $\mathcal{R}$. Both of these regions are assumed to be located at disjoint portions of the boundary $\mathcal{S}$. The rest of the boundary is assumed to be a forbidden-to-cross region $\mathcal{F}$ where no message should enter nor leave $\mathcal{D}$. The intensity $\sigma(x, y)$ of message generation on $\mathcal{O}$ is given, while the intensity $\rho(x, y)$ of received message on $\mathcal{R}$ is unknown. It is only assumed that these are consistent: the total flow of messages emitted and received are equal, *i.e.*,

$$\int_{\mathcal{R}} \rho(\mathbf{x}) \, ds = \int_{\mathcal{O}} \sigma(\mathbf{x}) \, ds \,. \tag{5.1}$$

The flow of messages is a vector field $\mathbf{T} : \mathcal{D} \to \mathbb{R}^2$, and $\varphi(\mathbf{x}) = \|\mathbf{T}(\mathbf{x})\|$ is its intensity. The congestion cost per packet transmitted at each point in $\mathcal{D}$ is a function $c(x, y, \varphi)$ of the point and the intensity φ of the flow of messages through that point.

We defined $\mathcal{Q} := \mathcal{O} \cup \mathcal{F}$ and extend the function σ to the whole of $\mathcal{Q}$ by setting $\sigma(\mathbf{x}) = 0$ on $\mathcal{F}$. The boundary conditions are given by

$$\forall \mathbf{x} \in \mathcal{Q} \quad \langle \mathbf{T}(\mathbf{x}), \mathbf{n}(\mathbf{x}) \rangle = -\sigma(\mathbf{x}). \tag{5.2}$$

There is no source nor sink of messages in $\mathcal{D}$, which such as in fluid models, implies the constraint

$$\forall \mathbf{x} \in \mathcal{D} \quad \nabla \cdot \mathbf{T}(\mathbf{x}) = 0\,. \tag{5.3}$$

which suffices to insure the consistency condition (5.1).

The path followed by a packet is specified by its direction of travel $e_\theta = (\cos\theta, \sin\theta)$ along its path, according to $\mathbf{x} = e_\theta$. The user-cost or the cost incurred by one packet traveling from $\mathbf{x}_0 \in \mathcal{O}$ at time t_0 to $\mathbf{x}_1 \in \mathcal{R}$ reached at time t_1 is

$$J(e_\theta(\cdot)) \;=\; \int_{\mathcal{S}} c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)\sqrt{dx^2 + dy^2} \tag{5.4}$$

$$= \int_{t_0}^{t_1} c(\mathbf{x}(t), \|\mathbf{T}(\mathbf{x}(t))\|)\; dt\,. \tag{5.5}$$

where $\mathcal{S}$ is the path such that $(x(t_0), y(t_0)) = \mathbf{x}_0$ and $(x(t_1), y(t_1)) = \mathbf{x}_1$ following at each time t the path given by $e_\theta(\cdot)$.

The system-cost is taken as

$$G(\mathbf{T}(\cdot)) = \int_{\mathcal{D}} c(\mathbf{x}, \|\mathbf{T}(\mathbf{x})\|)\|\mathbf{T}(\mathbf{x})\|\; d\mathbf{x}\,. \tag{5.6}$$

Quadratic congestion total cost

In this chapter, we focus on the quadratic congestion total cost that has been studied for the system-optimization problem in the work by Tassiulas and Toumpis [24]. As it was explained in Section 1.2, this election is justified by taken into account the work of Gupta and Kumar [45]. We assume in this context that the local cost is linear on the congestion, *i.e.*,

$$c(\mathbf{x}, \mathbf{T}) = \frac{1}{2}c(\mathbf{x})\|\mathbf{T}\|,$$

which implies that the total cost function for the system-optimization problem is

$$G(\mathbf{x}, \mathbf{T}) = \frac{1}{2}\int_{\mathcal{D}} c(x)\|\mathbf{T}\|^2\, dx.$$

The necessary optimality condition for the global optimization problem gives us that if we find p such that

$$\forall \mathbf{x} \in \mathcal{D} \quad \nabla \cdot \left(\frac{1}{c(\mathbf{x})}\nabla p(\mathbf{x})\right) = 0, \tag{5.7a}$$

$$\forall \mathbf{x} \in \mathcal{Q} \quad \frac{\partial p}{\partial n}(\mathbf{x}) = c(\mathbf{x})\sigma(\mathbf{x}), \tag{5.7b}$$

$$\forall \mathbf{x} \in \mathcal{R} \quad p(\mathbf{x}) = 0. \tag{5.7c}$$

Then p is optimal. We deduced that the optimal flow is given by

$$\mathbf{T}^*(\mathbf{x}) = \frac{\nabla p(\mathbf{x})}{c(\mathbf{x})}.$$

In the previous section we defined a continuum Wardrop equilibrium where each single message seeks its optimal solution to follow the path that minimizes the integral line of the cost function across the path, assuming that its lone deviation from that scheme would have no effect on the overall congestion map. In that setting the Wardrop equilibrium can be obtained by solving the system optimal problem in which the total cost density is replaced by $\int_0^{\varphi} c(\mathbf{x}, \varphi)\, d\varphi$. This implies that the system optimal solution and the Wardrop equilibrium coincide in the domain $\{\mathbf{x} \in \mathcal{D}; f^*(\mathbf{x}) \neq 0\}$.

In order to find a numerical solution to our problem we consider the Finite Element Method (FEM), largely used in numerical modeling of physical systems in disciplines such as Electromagnetism, Fluid Dynamics, and others.

The general variational problem is to seek $u \in V$ such that

$$(\mathrm{VP}) \begin{cases} a(u, v) = l(v) \\ \forall\, v \in V. \end{cases}$$

We seek the weak formulation of the set of equations (**??**). If we multiply equation (5.7a) by a generic function $\psi \in H^1_{\mathcal{R}}(\mathcal{D})$ where we consider the closed subspace

$$H^1_{\mathcal{R}}(\mathcal{D}) = \{v \in H^1(\mathcal{D});\, v(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in \mathcal{R}\},$$

and integrate over the domain $\mathcal{D}$, after a small manipulation using equations (5.7b) and (5.7c), we obtain that equation (5.7a) is equivalent to

$$\int_{\mathcal{D}} \frac{1}{c} \nabla p \cdot \nabla \psi \, dx = \int_{\mathcal{Q}} \sigma \cdot \psi \, dx.$$

The problem defined by the set of equations (**??**) is equivalent to the problem of seeking $p \in H^1_{\mathcal{R}}(\mathcal{D})$ such that

$$\int_{\mathcal{D}} \frac{1}{c} \nabla p \cdot \nabla \psi \, dx = \int_{\mathcal{Q}} \sigma \cdot \psi \, dx \quad \forall \psi \in H^1_{\mathcal{R}}(\mathcal{D}).$$

In that sense, if we consider the functions $a(\cdot,\cdot)$ and $l(\cdot)$ defined as

$$a(u, v) = \int_{\mathcal{D}} \frac{1}{c} \nabla u \cdot \nabla v \, dx \quad \text{and} \quad l(v) = \int_{\mathcal{Q}} \sigma \cdot v \, dx.$$

in the space $V = H^1_{\mathcal{R}}(\mathcal{D})$, we have set our problem as a variational problem, where the solution will be the function p.

In our case, the bilinear function $a(\cdot,\cdot)$ is V-elliptic, symmetric and continuous in $H^1(\mathcal{D})$ and the linear function $l(\cdot)$ is bounded. Then we can use Lions-Lax-Milgram Theorem and conclude that the solution exists and is unique.

This theorem gives us not only the existence and uniqueness of the solution but also gives us information about the stability of the solution when the data changes saying that the solution depends continuously on the data.

5.1.1 Why to use the Finite Element Method?

The idea of the Finite Element Method is to discretize the problem (VP) when the dimension of the space is infinite. This is interesting in our case as we are looking for a solution within the space of functions $H^1(\mathcal{D})$. From this approach we obtain a linear system for which there are many standard methods to solve it. An intern discretization of the variational problem is to take V_h as a vector subspace of finite dimension ($V_h \subseteq V$), where $h > 0$ is a discretization parameter such that when $h \to 0$, the dimension of V_h goes to infinity. Since $V_h \subseteq V$, $a(\cdot,\cdot)$, $l(\cdot)$ are well defined in V_h, then the discretized problem becomes to seek $u_h \in V_h$ such that

$$(EV_h) \left\{ \begin{array}{l} a(u_h, v_h) = l(v_h) \\ \forall v_h \in V_h, \quad \forall u_h \in V_h. \end{array} \right.$$

Due to Lions-Lax-Milgram we have a solution that depends continuously on the data.

Let $\{\varphi_1, \ldots, \varphi_{N_h}\}$ be a base of V_h, then we can write for every $u_h \in V_h$,

$$u_h = \sum_{j=1}^{N_h} \alpha_j \varphi_j,$$

where $\alpha = (\alpha_1, \ldots, \alpha_{N_h}) \in \mathbb{R}^{N_h}$ is unique for each $u_h \in V_h$. Then the equation of the (EV_h) problem for the base of V_h becomes

$$a\left(\sum_{j=1}^{N_h} \alpha_j \varphi_j, \varphi_i\right) = l(\varphi_i) \quad \forall i = 1, \ldots, N_h,$$

Then from the bilinearity and symmetry of $a(\cdot,\cdot)$, we obtain

$$\sum_{j=1}^{N_h} a(\varphi_i, \varphi_j)\alpha_j = l(\varphi_i) \quad i = 1, \ldots, N_h, \quad \alpha \in \mathbb{R}^{N_h},$$

which is equivalent to the linear system

$$A\alpha = b,$$

where $\alpha \in \mathbb{R}^{N_h}$, $A_{ij} = a(\varphi_j, \varphi_i)$, and $b_i = l(\varphi_i)$. In order to solve this linear system, we can use standard methods.

5.1.2 Description of the Finite Element Method

In this part we are going to give some definitions to understand the finite element method.

For every $k \geq 0$, we denote $\mathbb{P}_k$ to the *space of polynomial functions* from $\mathbb{R}^2$ to $\mathbb{R}$ of degree less or equal to k.

Definition 5.1.1 (2-simplex) *Consider* 3 *vertices*

$$a_1 = (a_{11}, a_{21}), \quad a_2 = (a_{12}, a_{22}), \quad a_3 = (a_{13}, a_{23}),$$

in $\mathbb{R}^2$*, not aligned (we call them non-degenerated vertices), then a* 2*-simplex* T *of vertices* $\{a_j\,;\, 1 \leq j \leq 3\}$ *is the convex hull of those vertices.*

As the vertices are not in the same line, every point $\mathbf{x} \in \mathbb{R}^2$ can be written as a linear combination of those vertices. We denote $\{\lambda_j(\mathbf{x})\,;\, 1 \leq j \leq 3\}$ the *barycentric coordinates* of the point $\mathbf{x}$ with respect to the vertices $\{a_j,\, 1 \leq j \leq 3\}$. Then, we can characterize the 2-simplex T of vertices $\{a_j, 1 \leq j \leq 3\}$ by

$$T = \left\{ x \in \mathbb{R}^2\,;\, x = \sum_{i=1}^{3} \lambda_j(x)\, a_j\,;\; 0 \leq \lambda_j(x) \leq 1\,,\, \forall\, 1 \leq j \leq 3 \right\}.$$

For the error analysis it is useful to consider the following geometric parameters:

$$\begin{aligned} h_T &= \text{diameter of } T \text{ (length of the greatest side)}, \\ \rho_T &= \text{roundness of } T \text{ (diameter of the greatest ball included in } T). \end{aligned}$$

We define the 2-simplex of reference $\hat{T}$ as the 2-simplex that has as vertices $\hat{a}_1 = (1,0), \hat{a}_2 = (0,1)$ and $\hat{a}_3 = (0,0)$. If we consider T as a non-degenerated 2-simplex of vertices $\{a_j, 1 \leq j \leq 3\}$, then there exists a unique invertible matrix $B_T \in \mathcal{M}_2(\mathbb{R})$ and a unique vector $b_T \in \mathbb{R}^2$ such that

$$\forall\, 1 \leq j \leq 3\,,\; a_j = B_T\, \hat{a}_j + b_T.$$

We denote $\mathcal{F}_T$ this affine transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$.

Construction of the finite elements

Let us consider:

- A compact set $T \subseteq \mathbb{R}^2$, connected and with non-empty interior.
- A finite set $\Sigma = \{a_j\}_{j=1}^{N}$ of N distinct points of T.
- A vectorial space P of finite dimension and made of functions from T to the set of real numbers.

Definition 5.1.2 (P-unisolvent) *The set Σ is called P-unisolvent, if and only if, given N scalars α_j, $1 \leq j \leq N$, there exists a function p on the space P and only one such that*

$$p(a_j) = \alpha_j, \quad 1 \leq j \leq N.$$

When the set Σ is P-unisolvent, the triplet (T, P, Σ) is called finite element of Lagrange.

Given these definitions we have that for every function v defined over T to real values, there exists a function $p \in P$ and only one, that interpolate v over Σ, *i.e.*, it satisfies

$$p(a_j) = v(a_j),\ 1 \leq j \leq N.$$

Definition 5.1.3 *Given a finite element of Lagrange (T, P, Σ), we call functions of base to the N functions p_i, $1 \leq i \leq N$, such that*

$$p_i(a_j) = \delta_{ij}, \quad 1 \leq j \leq N.$$

We call operator of P-interpolation of Lagrange over Σ to the operator that to any function v defined over T it gives

$$\Pi v = \sum_{i=1}^{N} v(a_i) p_i,$$

and Πv is called the P-interpolate of Lagrange of v over Σ.

Theorem 5.1.1 ([60],Theorem 4.5.) *Let $(\hat{T}, \hat{P}, \hat{\Sigma} = \{\hat{\varphi}_i; 1 \leq i \leq M\})$ be a finite element of Lagrange on $\mathbb{R}^2$ and let F be a bicontinuous bijective function from $\mathbb{R}^2$ to $\mathbb{R}^2$. Then the triplet $(T, P, \Sigma = \{\varphi_i; 1 \leq i \leq M\})$ defined by*

$$T = F(\hat{T}) \tag{5.8a}$$

$$P = \{\hat{p} \circ F^{-1}; \hat{p} \in \hat{P}\} \tag{5.8b}$$

$$dom(\Sigma) = \{v = \hat{v} \circ F^{-1}; \hat{v} \in dom(\hat{\Sigma})\} \tag{5.8c}$$

$$\forall v \in dom(\Sigma), \varphi_i(v) = \hat{\varphi}_i(\hat{v}), 1 \leq i \leq M, \tag{5.8d}$$

is also a finite element of Lagrange.

Definition 5.1.4 *Two finite elements of Lagrange $(\hat{T}, \hat{P}, \hat{\Sigma} = \{\hat{\varphi}_i; 1 \leq i \leq M\})$ and $(T, P, \Sigma = \{\varphi_i; 1 \leq i \leq M\})$ are equivalents if there exists a bicontinuous bijective function F from $\hat{T}$ to T such that (T, P, Σ) satisfy (5.8). If F is an affine transformation, we say that they are affine-equivalents.*

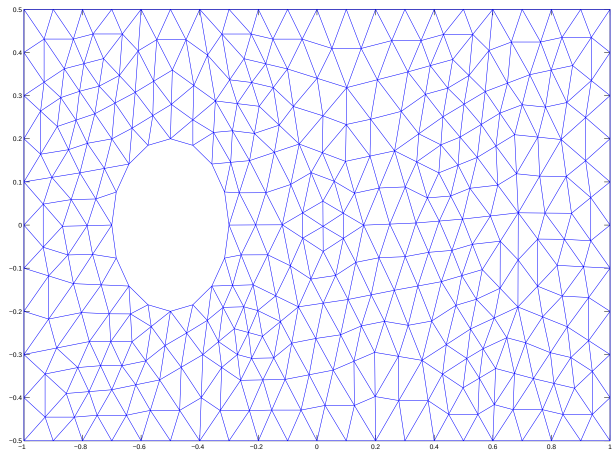


Figure 5.1: Triangulation of the domain $[-1,1] \times [-0.5,0.5] \backslash D((-0.5,0),0.2)$.

From the elements described above we realize that we can construct from the 2-simplex of reference $\hat{T}$ all the 2-simplex that we want and that these 2-simplex will be affine-equivalents to the 2-simplex of reference.

Suppose the domain $\mathcal{D}$ has a polygonal boundary $\mathcal{S}$, then we can cover the closure of $\mathcal{D}$ (the union of $\mathcal{D}$ and $\mathcal{S}$, denoted by $\bar{\mathcal{D}}$) by a triangulation, *i.e.*,

$$\bar{\mathcal{D}} = \bigcup_{T \in \mathcal{T}} T$$

and each T is a closed triangle where

- Every $T \in \mathcal{T}$ is a triangle.
- The interior of two different triangles are disjoint.
- Every face of a triangle is either the face of another triangle (in which case, they are called adjacent) or a part of the boundary.

As an example, in Figure 5.1 we have a rectangular domain of boundary $[-1,1] \times [-1,1]$ minus the disk of center $(-0.5,0)$ and radius 0.2, and triangulation approximation defined on this domain.

For convention, $\mathcal{T}_h$ denotes a triangulation of $\bar{\mathcal{D}}$ such that

$$h = \max_{T \in \mathcal{T}_h} h_T \,,$$

where h_T is the diameter of the polygon T.

Suppose that for every polygon T of $\mathcal{T}_h$, there is associated a finite element of Lagrange (T, P_T, Σ) such that

$$P_T \subseteq H^1(T) \,,$$

and we define the finite dimensional spaces

$$X_h = \{v \in \mathcal{C}^0(\bar{\mathcal{D}})\,;\, \forall\, T \in \mathcal{T}_h,\, v|_T \in P_T\} \tag{5.9}$$
$$X_{0h} = \{v \in X_h\,;\, v|_{\mathcal{R}} = 0\}. \tag{5.10}$$

If (T, P, Σ) is a finite element of Lagrange, to any function v defined over T, we associate the function Πv that P-interpolates v over T. The idea is to study an upper bound for the interpolation error $v - \Pi v$ with the norm $H^1(T)$.

Let T be a compact of $\mathbb{R}^2$, connected and with non-empty interior. For simplicity we denote $H^m(T)$ at the Sobolev space $H^m(\mathring{T})$ where $\mathring{T}$ is the interior of T.

Definition 5.1.5 *Let T be a polygon on $\mathbb{R}^2$. A finite element (T, P, Σ) is called of class $\mathcal{C}^0$ if the two following conditions are satisfied:*

1. *$P \subseteq \mathcal{C}^0(T)$,*
2. *For every face T' of T, the set $\Sigma' = \Sigma \cap T'$ is P'-unisolvent where $P' = \{p|_{T'};\, p \in P\}$*

Definition 5.1.6 *We call $(\mathcal{T}_h)$ a family of regular triangulations of $\bar{\mathcal{D}}$ if the following four conditions are satisfied:*

1. *All the finite elements (T, P_T, Σ_T) of every triangulation are affine-equivalents to the same finite element of reference $(\hat{T}, \hat{P}, \hat{\Sigma})$ of class $\mathcal{C}^0$.*
2. *For every pair $(\hat{T}'_1, \hat{T}'_2)$ of faces of $\hat{T}$ and for every application $\hat{F}$ affine invertible and from $\mathbb{R}^2$ to $\mathbb{R}^2$ such that $\hat{T}'_2 = \hat{F}(\hat{T}'_1)$, we have*
$$\hat{\sigma} \cap \hat{T}'_2 = \hat{F}(\hat{\sigma} \cap \hat{T}'_1)$$
and
$$\{\hat{p}|_{K'_2};\, \hat{p} \in \hat{P}\} = \{p \circ \hat{F}|_{K'_1};\, p \in \hat{P}\}.$$
3. *We have*
$$h = \max_{T \in \mathcal{T}_h} h_T \to 0.$$
4. *There exists a constant $\sigma \geq 1$ such that*
$$\forall h,\, \forall T \in \mathcal{T}_h,\, \frac{h_T}{\rho_T} \leq \sigma.$$

Theorem 5.1.2 *Let $\mathcal{D}$ be an open polytope of $\mathbb{R}^d$, $d \leq 3$. Let $(\mathcal{T}_h)$ be a family of regular triangulations of $\bar{\mathcal{D}}$ associated to a finite element of reference $(\hat{T}, \hat{P}, \hat{\Sigma})$ of class $\mathcal{C}^0$. We suppose that there exists an integer $k \geq 1$ such that*

$$P_k \subseteq \hat{P} \subseteq H^1(\hat{K})$$

Then the finite element method is convergent, i.e. the solution u_h of the problem (VP$_\mathrm{h}$) *converges to the solution of* (VP) *in $H^1(\mathcal{D})$:*

$$\lim_{h\to 0}\|u-u_h\|_{1,\mathcal{D}}=0.$$

There exists a constant C independent of h such that if the solution belongs to the Sobolev space $H^{k+1}(\mathcal{D})$

$$\|u-u_h\|_{1,\mathcal{D}}\leq Ch^k\|u\|_{k+1,\mathcal{D}}.$$

5.2 Examples of applications

1.- Gathering in Wireless Sensor Networks

Suppose that we want to protect a land for growing crops from an external threat such as forest fire. We deploy uniformly a large quantity of sensor nodes over the land and we want to give a description of the flow according to this setting and a description of the arrival of the new information that is coming from the sensor nodes which are on the boundary of the land. As we are considering an external threat we consider that only the external nodes are going to generate information and the interior nodes are going to serve only as relay nodes.

We suppose that the generation of information can be approximated by $\sigma=1$ for every sensor node in the boundary. We also consider that the center of analysis of information is located inside the domain and for simplicity we suppose that it can be modeled as a closed set with non-empty interior.

We suppose that our land can be modeled by the rectangle $[-1,1]\times[-0.5,0.5]$ and the center of analysis of information is located at the point $(-0.5,0)$. If we suppose that we distribute uniform over the whole network, then it is reasonable to assume that if the land doesn't have other environmental problems the cost of the network will be uniform and for simplicity we consider $c=1$.

After using the model explained in the previous section the direction of the flow of information can be described by the red lines in Figure 5.2.

2.- Dafermos Example

We consider a rectangular domain $[-1,1]\times[-0.5,0.5]$. Following the paper of Dafermos [29] we impose that the value of the flow in the boundary is uniform and equal to 1 in the vertical left boundary and in the horizontal lower boundary.

We suppose that we distribution of the nodes inside the domain is uniform, then it is reasonable to assume that if the land doesn't have other environmental problems, the transmission cost over the network is uniform and for simplicity we consider $c=1$.

Then the description of the flow of information is given by the red lines in Figure 5.3.

3.- Example with obstacles

Once more, for simplicity we consider a rectangular domain but this time for problems in the land we consider that some of the relay nodes can not be put in a specific area. We

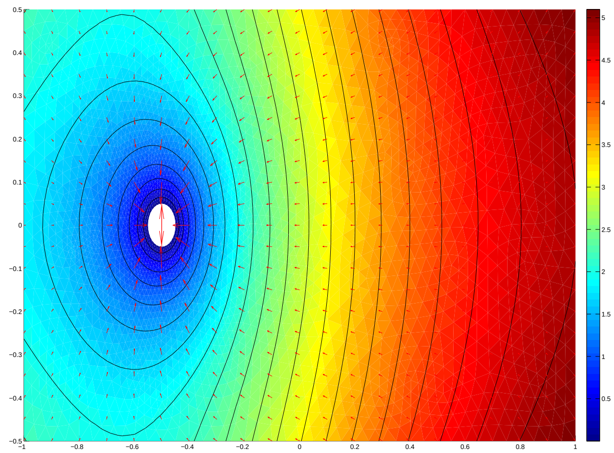


Figure 5.2: Solution for a wireless sensor network.

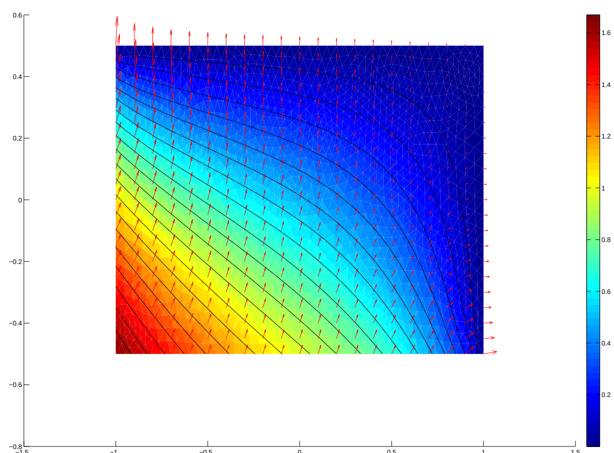


Figure 5.3: Problem similar to Dafermos' problem.

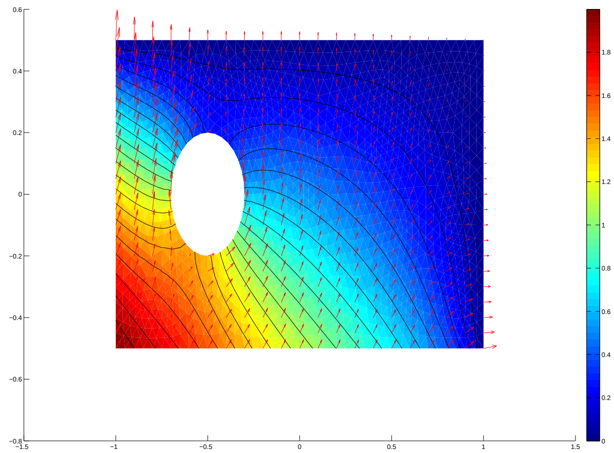


Figure 5.4: When the domain considered is not simply connected.

suppose that the distribution of the nodes is uniform and we suppose that the cost of the network will be uniform $c = 1$.

We consider that the domain is give by $[-1, 1] \times [-0.5, 0.5]$ and the hole is modeled as a circle with center in $(-0.5, 0)$ and radius 0.2.

Then the direction of the flow of information is given by the red lines in Figure 5.4.

5.3 Conclusions

In the present work we have used the Finite Element Method in solving the routing problem in massively dense static ad-hoc networks. The node density in these massively dense systems is approximated by a continuous area with costs depending on the location and the congestion of the network.

The problem considered is that messages have to flow from a region $\mathcal{O}$ of the boundary $\mathcal{S}$ of a domain $\mathcal{D}$ to a disjoint region $\mathcal{R}$ of $\mathcal{S}$. The intensity of message generation on $\mathcal{O}$ is given. In this framework we study the case of linear congestion cost per packet. We mention a result from [61] on existence and uniqueness of the solution and present the stability of this solution with respect to the initial flow.

Numerically we obtain via the Finite Element Method an approximation of the solution and prove a result of convergence and the velocity of convergence of this numerical approximation to the exact solution.

Notations

Some of the notations used in this chapter are currently used in the Partial Differential Equations (PDE) community, but may not be well known in our community.

The functions in $L^1_{\text{loc}}(\mathcal{D})$, also called locally integrable functions, are the functions which are integrable on any compact set of its domain of definition $\mathcal{D}$.

The functions in $L^2(\mathcal{D})$ are the functions that are square integrable, *i.e.*

$$L^2(\mathcal{D}) = \left\{ f : \mathcal{D} \to \mathbb{R}; \int_{\mathcal{D}} \|f(x)\|^2 \, dx < +\infty \right\}.$$

Note: The functions in $L^2(\mathcal{D})$, also belong to $L^1_{\text{loc}}(\mathcal{D})$.

The functions in $\mathcal{C}_c(\mathcal{D})$ are the continuous functions with compact support over $\mathcal{D}$. The functions in $\mathcal{C}^k(\mathcal{D})$ are the functions k times continuously differentiable over $\mathcal{D}$ $(k \geq 0)$.

The function in $\mathcal{C}^\infty(\mathcal{D}) = \bigcap\limits_{k \geq 0} \mathcal{C}^k$ are the functions for which all its derivatives are differentiable over $\mathcal{D}$.

The function in $\mathcal{C}_c^\infty = \mathcal{C}^\infty \cap \mathcal{C}_c(\mathcal{D})$ are the functions with compact support over $\mathcal{D}$ for which all its derivatives are differentiable over $\mathcal{D}$.

Given a function $u \in \mathcal{C}^1(\mathcal{D})$ and a function $\varphi \in \mathcal{C}_c^\infty$, using the integration by parts formula we obtain

$$\int_{\mathcal{D}} u \frac{\partial \varphi}{\partial x_i} \, dx = - \int_{\mathcal{D}} \frac{\partial u}{\partial x_i} \varphi \, dx \quad \forall i = 1, 2. \tag{5.11}$$

There are no boundary terms, since φ has compact support in $\mathcal{D}$ and thus vanishes near the boundary $\partial \mathcal{D}$. The left hand side of equation (5.11) makes sense even if $u \in L^1_{\text{loc}}(\mathcal{D})$. However, the expression $\frac{\partial u}{\partial x_i}$ on the right hand side doesn't have a meaning. Then in Partial Differential Equations people work with a concept called weak partial derivative.

Definition 5.3.1 *Suppose $u, v \in L^1_{\text{loc}}(\mathcal{D})$. We say that v is the weak partial derivative of u, written $v = Du$ provided*

$$\int_{\mathcal{D}} u \frac{\partial \varphi}{\partial x_i} = - \int_{\mathcal{D}} v_i \varphi \quad \forall \varphi \in \mathcal{C}_c^\infty(\mathcal{D}) \quad \forall i = 1, 2.$$

The functions in $H^1(\mathcal{D})$ are the functions that are square integrable and whose weak partial derivative is also square integrable.

$$H^1(\mathcal{D}) = \left\{ u \in L^2(\mathcal{D}); Du \in L^2(\mathcal{D}) \right.$$

For more information about the motivation to work on these spaces consult chapter VIII and IX of the book on Functional Analysis [62].

Chapter 6

Magnetworks: Mobility of the Nodes

In this chapter, we modelize and analyze mobile ad hoc networks. The mobility of the nodes in this type of network is not a negligible fact. Recall that to determine, at each stage of the route, to which node forward data is made dynamically based on the network connectivity. Consequently, in the presence of mobility, traditional routing schemes meant for wired networks are not appropriate for a mobile ad hoc environment. Because of this, one of the most challenging problems in the performance analysis of this type of networks has been the routing problem.

6.1 Introduction

In this chapter, we are mainly interested on the mobility issue on mobile ad hoc networks. In recent years, the idea of designing wireless networks where mobile terminals would themselves serve as relay nodes and route communications in a completely descentralized and self-organized fashion has generated a lot of interest. The growth of laptops and 802.11 wireless networking have made mobile ad hoc networks an increasingly popular research topic since the mid- to late 1990s, for the potential possibility to communicate between mobile terminals without the need of access points or base stations. There has been a particular interest in the application of this type of network in other situations. For example:

- Vehicular Ad Hoc Networks (VANETs), used for communication among vehicles and between vehicles and roadside equipment. This could give safety and comfort for passengers by providing collision warnings, road sign alarms and in-place traffic view to decide the best path along the road.

- Disruption-Tolerant Networking (DTN) that address the technical issues involving heterogeneous networks that may lack continuous network connectivity. The disruption may occur because of the limits of wireless radio range, energy resources, attack, and noise. One of many applications could be to upgrade the operating system of laptops or mobile phones without ever connecting to the Internet by simply being near another mobile terminal who has an updated version.

There has been as well industrial applications for mobile ad hoc networks. For example, the Swedish company TerraNet AB presented in 2007 a mesh network of mobile phones that allowed calls and data to be routed between participating handsets, without the need for a mobile phone base station [63]. Another example is the US non profit organization One Laptop per Child (OLPC) program. This program make use of IEEE 802.11s based ad hoc wireless mesh networking chip to create an affordable educational device for use in the developing world where the set of laptops establish a mobile ad hoc network [64]. In order to analyze the continuous approximation problem for this type of network, we first need to redefine some terms used in the context of static wireless ad hoc networks to mobile ad hoc networks.

In mobile ad hoc networks, the information density function ρ, the traffic flow function $\mathbf{T}$, and the node density function η, previously defined, may depend on time, *i.e.*, $\rho = \rho(\mathbf{x}, t)$, $\mathbf{T} = \mathbf{T}(\mathbf{x}, t)$, and $\eta = \eta(\mathbf{x}, t)$. We consider the routing problem within a time window $t \in [t_i, t_f)$ where t_i is the initial time and t_f is the final time. By analogy with Electromagnetism, we define the continuous *node current* $\mathbf{J}$ as the density of nodes $\rho(\mathbf{x}, t)$ in the position $\mathbf{x}$ multiplied by the nodes average drift velocity $\mathbf{v}(\mathbf{x}, t)$, *i.e.*,

$$\mathbf{J} = \rho(\mathbf{x}, t)\, \mathbf{v}(\mathbf{x}, t).$$

The rate at which nodes leaves an area (or volume) V, bounded by a curve (or surface) S, is given by the following expression:

$$\oint_S \mathbf{J} \cdot dS. \tag{6.1}$$

Since the information density function is conserved in the plane this integral must be equal to

$$\oint_S \mathbf{J} \cdot dS = -\frac{d}{dt} \oint_S \rho \cdot \mathbf{n}\, dS = -\int_V \frac{\partial \rho}{\partial t}\, dV. \tag{6.2}$$

From the divergence theorem and imposing the equality between the equations (6.1) and (6.2), we obtain the equivalent to Kirchhoff's current law:

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0.$$

Notice as well that

$$\nabla \cdot \mathbf{J} = \nabla \cdot (\rho\, \mathbf{v}) = \mathbf{v} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{v}.$$

We assume that we know the initial distribution of the sources and the destinations at time 0 denoted by ρ_0. Thus we obtain the following transport equation (TE):

$$(\mathrm{TE}) \begin{cases} \frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{v} = 0 & \text{in } D \times (0, T) \\ \rho(0) = \rho_0 & \text{on } D \times \{0\}. \end{cases}$$

The previous system of equations is known as the linear transport equation with initial condition for which a known solution (see Proposition II.1 of [65]) exists. A solution can be easily found in numerical softwares such as Matlab (PDE toolbox), Octave (PDE toolbox), or others.

Notice that given the initial distribution of the sources and destinations and the velocity of these distributions, we are able to compute the evolution of sources and destinations distributions over time $t \in [t_i, t_f)$. The velocity of sources and destinations may be estimated by having some previous knowledge about the behavior of these sources and destinations in our network such as, for example, measuring day-to-day variability in its travel behavior using GPS data, or by statistical inference.

In this chapter, our objective is to minimize the number of relay nodes $N(D)$ in the grid area network D needed to support the information created by the distribution of sources and received by the distribution of destinations. This should be done subject to the flow conservation condition, and knowing that the distribution of mobile sources and mobile destinations is the solution to the system of equations (TE). Thus our problem reads for all $t \in [t_i, t_f)$

$$\text{Min } N(D,t) = \int_D \eta(\mathbf{x},t)\, d\mathbf{x} = \int_D \|\mathbf{T}(\mathbf{x},t)\|^2\, d\mathbf{x} \tag{6.3}$$

$$\text{subject to} \quad \nabla \cdot \mathbf{T}(\mathbf{x},t) = \rho(\mathbf{x},t) \text{ in } D, \tag{6.4a}$$

$$\mathbf{T} \cdot \mathbf{n} = 0 \text{ on } S. \tag{6.4b}$$

where $\rho(\mathbf{x},t)$ is the solution to the problem (TE).

We recall that Tassiulas and Toumpis proved in [48] that among all traffic flow functions that satisfy equation (6.4a), the one that minimizes the number of nodes needed to support the network, must satisfy

$$\nabla \times \mathbf{T} = 0. \tag{6.5}$$

Using Helmholtz's theorem (also known as fundamental theorem of vector calculus) to last equation (6.5) we conclude that there exists a scalar potential function φ such that

$$-\nabla\varphi = \mathbf{T}. \tag{6.6}$$

Replacing this function into the conservation equation (6.4a), we obtain that

$$-\Delta\varphi = \rho \quad \text{for all } t \in [t_i, t_f). \tag{6.7}$$

We impose that no information is leaving the considered region D in mean. This implies that we are not considering the case where some of the nodes may leave the region. In equation (6.4b) from equation (6.6), this condition translates into $\nabla\varphi \cdot \mathbf{n} = 0$. From equation (6.7) and the last condition, we obtain the following system

$$\text{(LE)} \begin{cases} -\Delta\varphi = \rho & \text{in } D, \\ \nabla\varphi \cdot \mathbf{n} = 0 & \text{on } S, \end{cases} \tag{6.8}$$

which is the *Laplace equation* with Neumann boundary conditions.

If the function f is square integrable then the Laplace equation with Neumann boundary conditions has a unique solution[1] in $H^1(D)/\mathbb{R}$.

[1] As stated in the previous chapter there are many cases where the uniqueness of the solution to a transportation problem does not hold.

In summary, in order to solve our problem given by the equations (6.3), (6.4a), (6.4b), and (TE) we need to:

1. Solve the system of equations (TE),
2. Put the solution as input into the system of equations (LE),
3. Solve the system of equations (LE).

In the following, we give an example of this resolution where you can get explicit solutions.

Example 6.1.1 *We consider the one-dimensional case with $t_i = 0$ and $t_f = 2$ hours. We consider an initial distribution of sources (denoted ρ_0^+) and an initial distribution of destinations (denoted ρ_0^-) on the positive real line $[0, +\infty)$. We scale these distributions to be probability distributions. They represent in each location the proportion of sources and the proportion of destinations:*

$$\rho_0^+ = k_1 e^{-(x-3)^2} \quad \textit{and} \quad \rho_0^- = -k_2 e^{-(x-10)^2},$$

where k_1 and k_2 are normalization factors given by $k_1 = \frac{2}{\sqrt{\pi}\mathrm{erfc}(-3)}$, $k_2 = \frac{2}{\sqrt{\pi}\mathrm{erfc}(-10)}$, and erfc(x) *is the complementary error function defined as* $\mathrm{erfc(x)} = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-s^2}\, ds$.

We consider that the nodes average drift velocity is given by $v(x,t) = x$. We can think, as an example, of a highway where the cars are equipped with sensors and while they are advancing on the highway they can go faster as they advance.

1.- We first need to solve the transportation equation system (TE)*, that in our example reads*

$$\begin{cases} \frac{\partial \rho}{\partial t} + \frac{\partial (x\rho)}{\partial x} = 0 & \textit{on } \mathbb{R}_+ \times [0,T), \\ \rho(0) = \rho_0^+ + \rho_0^- & \textit{on } \mathbb{R}_+. \end{cases}$$

Using the method of characteristics we obtain that the information density function over time is given by $\rho(x,t) = \rho^+(x,t) + \rho^-(x,t)$ where

$$\begin{cases} \rho^+(x,t) = k_1 e^{-(xe^{-t}-3)^2 - t}, \\ \rho^-(x,t) = -k_2 e^{-(xe^{-t}-10)^2 - t}. \end{cases}$$

The solution combining the sources and destinations information density function over time are showed in Fig. 6.1.

2.- We use the above solution as input into the system of equations (LE). From the conservation equation we obtain

$$\frac{\partial T(x,t)}{\partial x} = \frac{\partial T^+(x,t)}{\partial x} + \frac{\partial T^-(x,t)}{\partial x}$$

$$\textit{where} \quad \frac{\partial T^+(x,t)}{\partial x} = \rho^+ \quad \textit{and} \quad \frac{\partial T^-(x,t)}{\partial x} = -\rho^-,$$

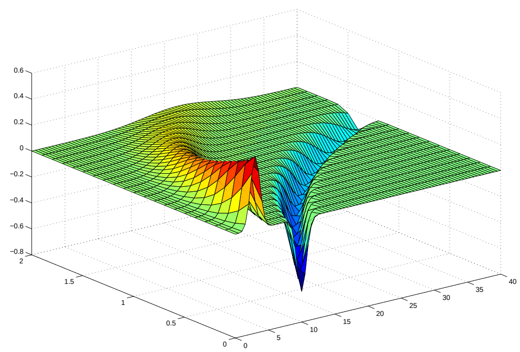


Figure 6.1: Distribution of the sources and destinations in the same line over time $[0, 2]$ hours. The initial proportion of sources and destinations is given by $\rho_0^+ = k_1 e^{-(x-3)^2}$ and $\rho_0^- = -k_2 e^{-(x-10)^2}$, with average drift velocity $v(x,t) = x$.

with initial condition that the flow is zero at the boundary point zero, i.e. $T(0,t) = 0$.

3.- We solve the Laplacian system of equations: Then the optimal traffic flow function is given by $T^*(x,t) = T^+(x,t) + T^-(x,t)$ *where*

$$\begin{cases} T^+(x,t) = \int_0^x k_1 e^{-(xe^{-t}-3)^2-t}\,dx \\ T^-(x,t) = -\int_0^x k_2 e^{-(xe^{-t}-10)^2-t}\,dx. \end{cases}$$

Thus the minimal number of active relay nodes needed to support the optimal flow at every time t *will be given by*

$$N^*(t) = \int_0^{+\infty} \|T^*(x,t)\|^2\,dx$$

which can be solved numerically.

In next section we present another type of mobility model where we consider the randomness in the mobility of the users.

6.2 Brownian mobility model

One of the most used mobility models used in networks is the Random Walk Mobility Model also known as the Brownian Mobility Model (see the survey [66] and the references therein).

If we have previous knowledge about the velocity drift of the distribution of information created at the sources (denoted ρ^+) and/or the distribution of information received at the destinations (denoted ρ^-), and we assume the Brownian mobility model, then the distribution of sources and/or the distribution of the destinations evolves according to the stochastic differential equation

$$d\rho^+(t) = \mathbf{v}^+(\mathbf{x},t)\,dt + \sigma_+(\mathbf{x},t)\,dW^+(t)$$

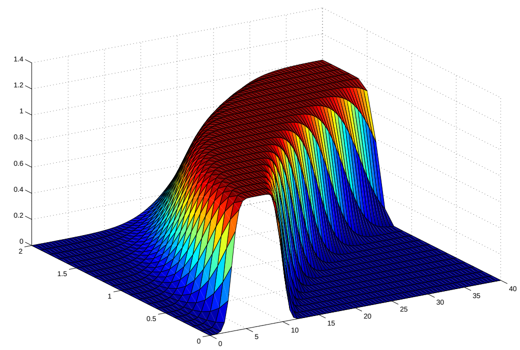


Figure 6.2: Optimal traffic flow over time $[0, 2]$ hours. The initial proportion of sources and destinations is given by $\rho_0^+ = k_1 e^{-(x-3)^2}$ and $\rho_0^- = -k_2 e^{-(x-10)^2}$, with average drift velocity $v(x, t) = x$.

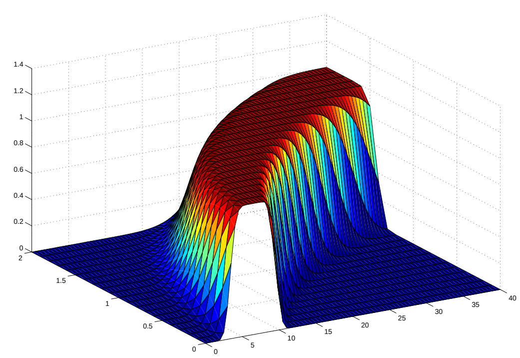


Figure 6.3: Optimal relay node distribution over time $[0, 2]$ hours. The initial proportion of sources and destinations is given by $\rho_0^+ = k_1 e^{-(x-3)^2}$ and $\rho_0^- = -k_2 e^{-(x-10)^2}$, with average drift velocity $v(x, t) = x$.

$$\text{and/or} \quad d\rho^-(t) = \mathbf{v}^-(\mathbf{x},t)\,dt + \sigma_-(\mathbf{x},t)\,dW^-(t),$$

where $W^+(t)$ and $W^-(t)$ are two independent Brownian motions with values in $X \times Y$ and the variance of the Brownian motions for sources and destinations, $\sigma_+ := \sigma_+(\mathbf{x},t)$ and $\sigma_- := \sigma_-(\mathbf{x},t)$, are parameters of the model.

Assume as in the previous case that we know the initial distribution of the information created at the sources. Then by using Itô's lemma [67], ρ^+ evolves in time by the Kolmogorov Forward Equation

$$\frac{\partial}{\partial s}p(\mathbf{x},s) = -\frac{\partial}{\partial x}[\mathbf{v}^+(\mathbf{x},s)p(\mathbf{x},s)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma_+^2 p(\mathbf{x},s)],$$

for $s \geq 0$, with initial condition $p(\mathbf{x},0) = \rho^+(\mathbf{x})$. Equivalently, the initial distribution of the destinations evolves in time by the Kolmogorov Forward Equation

$$\frac{\partial}{\partial s}p(\mathbf{x},s) = -\frac{\partial}{\partial x}[\mathbf{v}^-(\mathbf{x},s)p(\mathbf{x},s)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma_-^2 p(\mathbf{x},s)].$$

for $s \in [t_i, t_f)$, with initial condition $p(x,t_i) = \rho^-(x)$.

6.3 Optimization over time

Notice that the minimization problem we solve does not really consider the interaction on time because the problem describes the motion of sources and destinations nodes in the space and then we solve the static problem at each time.

The problem solved at each time may not be optimal in the whole period of time considered. Another more realistic problem would be to minimize the quantity of nodes used in the whole network during a fixed period of time $[t_i, t_f]$, *i.e.*,

$$\begin{aligned}\text{Min} \int_{t_i}^{t_f} N(D,t)\,dt &= \int_{t_i}^{t_f}\int_D \eta(\mathbf{x},t)\,d\mathbf{x} = \\ &= \int_{t_i}^{t_f}\int_D \|\mathbf{T}(\mathbf{x},t)\|^2\,d\mathbf{x}\end{aligned}$$

where $N(D,t)$ is the number of active relay nodes in the network D at time t, subject to (6.4a), (6.4b), and (TE).

For the case that we have randomness in the system, this problem is given as

$$\text{Min} \int_{t_i}^{t_f} \mathbb{E}\{N(D,t)\}\,dt = \int_{t_i}^{t_f}\int_D \mathbb{E}\left\{\|\mathbf{T}(\mathbf{x},t)\|^2\right\}\,d\mathbf{x} \tag{6.9}$$

subject to (6.4a), (6.4b), and (TE) since D is compact. From the work of Santambrogio ([68], page 6) we have the following result: The problem

$$\text{Min} \int_D k(\mathbf{x})\|\mathbf{T}(\mathbf{x})\|\,d\mathbf{x} \quad \text{such that} \quad \nabla\cdot\mathbf{T} = \mu - \nu,$$

is equivalent by duality to the problem of finding

$$\text{Min} \int_{D\times D} d_k(x,y)\, d\gamma \quad \text{such that} \quad \gamma \in \Pi(\mu,\nu) \quad \text{where}$$

$$d_k(x,y) = \inf_{\substack{\{\omega\,:\,\omega(0)=x,\\ \omega(1)=y\}}} L_k(\omega) := \int_0^1 k(\omega(t))\|\omega'(t)\|\, dt.$$

and $\Pi(\mu,\nu)$ is the set of probability distributions with marginals μ and ν. In our case $k(\mathbf{x}) = \|\mathbf{T}(\mathbf{x})\|$ such that

$$L_k(\omega) = \int_0^1 \|\mathbf{T}(\omega(t))\|\|\omega'(t)\|\, dt.$$

Given that $\omega(0) = x$, and $\omega(1) = y$ then by change of variables $L_k(\omega) = \int_x^y \|\mathbf{T}(\mathbf{x})\|\, d\mathbf{x}$, and as it is independent of ω then $d_k(x,y) = \int_x^y \|\mathbf{T}(\mathbf{x})\|\, d\mathbf{x}$.

Example 6.3.1 *For the case where we do not have previous knowledge about the velocity drift then we just consider the standard Brownian mobility model given by*

$$d\rho^+(t) = \sigma_+(\mathbf{x},t)\, dW^+(t),$$

and/or

$$d\rho^-(t) = \sigma_-(\mathbf{x},t)\, dW^-(t),$$

where $W^+(t)$ and $W^-(t)$ are two independent Brownian motions with values in $X \times Y$.

Then, the previous equations translate into

$$\frac{\partial}{\partial s}p(x,s) = +\frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma_+^2(x,s)p(x,s)],$$

$$\frac{\partial}{\partial s}p(x,s) = +\frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma_-^2(x,s)p(x,s)],$$

which have as solution the following equations

$$\rho^+(x,t) = \frac{1}{\sqrt{2\pi t\sigma^+}}e^{-\frac{x^2}{2t\sigma^+}},$$

$$\rho^-(x,t) = \frac{1}{\sqrt{2\pi t\sigma^-}}e^{-\frac{x^2}{2t\sigma^-}}.$$

Now we can replace this solution into Step 2 and Step 3.

Remark 6.3.1 *Notice that if we suppose that the distribution of the destinations is fixed, as it will be the case for aggregation centers of information, then $\sigma^- = 0$ and then $\rho^-(x,t) = \rho^-$ for all time t.*

Part II

Optimal Planning of Massively Dense Cellular Networks

Chapter 7

Capacity of Massively Dense Networks with MIMO capabilities

7.1 Introduction

More than sixty years ago, Shannon [69] provided a mathematical framework to analyze fundamental limits of information transfer for the case of single-input and single-output channels. He introduced the *channel capacity* as the maximum rate at which information can be reliably transmitted through the channel. From a purely theoretical point of view, there is no bound on the capacity as both bandwidth and power can be arbitrarily high. However, in practice, we can only transmit with finite power and over a restricted frequency band for physical and regulatory reasons. Recently, multiple-input and multiple-output (MIMO) systems have been extensively studied since significant growth in terms of capacity has been predicted (see *e.g.* [70, 71]). More specifically, in a system with n_{T} transmit and n_{R} receive antennas the capacity scales linearly with $\min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$ for independent and identically distributed (i.i.d.) Gaussian channels, at high signal-to-noise ratio (SNR). Again, MIMO systems suggest that the capacity can increase to infinity if the number of antennas grows large at both the transmitter and the receiver.

One possible alternative to achieve the capacity of MIMO systems from conventional SISO systems is through cooperation between mobile devices. These systems are known as network MIMO systems, distributed MIMO systems, or virtual antenna array systems. Mobile devices use the partnered mobile device's antenna as virtual antennas.

Recent works [72] have shown that the capacity, even for an increasing number of (virtual) antennas, is limited by the density of scatterers in the environment. In other words, the number of antennas should be less than the number of degrees of freedom (modes) provided by the channel. The goal of this chapter is to show that, even when the channel offers an infinite number of modes, the capacity is mainly limited by the ratio between the size of the antenna array at the base station and at the mobile terminal and the wavelength, which we call the *space frontier*. Indeed, in general, for a given space, increasing n_{T} or n_{R} decreases the relative distances between the antennas. Once the distance is less than half the transmit signal wavelength λ the antennas become correlated [73] and the capacity does not grow linearly anymore. In case of a circular antenna array, it has been demonstrated by Pollock [74] that the capacity saturates if the number of antennas increases. In this work, we aim to extend Pollock's contribution to one- and two-dimensional antenna arrays. We study the capacity limits of Network MIMO channels as well as of MIMO Gaussian broadcast channels (MIMO-GBC) with linear precoding. In the latter, we assume a single transmitter modeled as a dense line of antennas which transmits to many independent single-antenna receivers. The general capacity solutions for those schemes are mathematically involved [75] and require the application of results from random matrix theory and free probability [76].

7.2 Random matrix theory tools

Since the pioneering work of Wigner [77] on the asymptotic empirical eigenvalue distribution of random Hermitian matrices, random matrix theory has grown into a new field of research in theoretical physics and applied probability. The main application to communications lies in the derivation of asymptotic results for large matrices. Specifically, the eigenvalue distribution of large Hermitian matrices converges, in many practical cases, to a definite probability distribution, called *empirical distribution*. For instance, if $\mathbf{X} \in \mathcal{M}(\mathbb{C}, N, L)$ is a $N \times K$ *Gaussian* matrix (*i.e.* a matrix with Gaussian i.i.d. entries), the eigenvalue distribution of the matrix $\frac{1}{L}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is known to converge, when $N, L \to \infty$ and $N/L \to c$, towards the *Marčhenko-Pastur law* μ_c [76]. Random Matrix Theory provides many tools to handle the empirical distribution of large random matrices. Among those tools, the *Stieltjes transform* $\mathcal{S}_{\mathbf{X}}$ of a large Hermitian matrix $\mathbf{X}$, defined on the half complex space $\{z \in \mathbb{C}, \mathrm{Im}(z) > 0\}$, is

$$\mathcal{S}_{\mathbf{X}}(z) = \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} f(\lambda) \mathrm{d}\lambda \tag{7.1}$$

where f is the empirical distribution of eigenvalues of $\mathbf{X}$.

Silverstein [78] derived a fixed-point expression of the Stieltjes transform for a particular random matrix structure in the following theorem,

Theorem 7.2.1 *Let the entries of the $N \times K$ matrix $\mathbf{W}$ be i.i.d. with zero mean and variance $1/N$. Let $\mathbf{X}$ be an $N \times N$ Hermitian random matrix with an empirical eigenvalue distribution function converging weakly to $P_{\mathbf{X}}(x)$ almost surely. Moreover, let $\mathbf{Y}$ be a $K \times K$ real diagonal random matrix with an empirical distribution function converging almost surely in distribution to a probability distribution function $P_{\mathbf{Y}}(x)$ as $K \to \infty$. Then almost surely, the empirical eigenvalue distribution of the random matrix:*

$$\mathbf{H} = \mathbf{X} + \mathbf{W}\mathbf{Y}\mathbf{W}^{\mathsf{H}} \tag{7.2}$$

converges weakly, as $K, N \to \infty$ but $K/N \to \alpha$ fixed, to the unique distribution function whose Stieltjes transform satisfies:

$$\mathcal{S}_{\mathbf{H}}(z) = \mathcal{S}_{\mathbf{X}}\left(z - \alpha \int \frac{y}{1 + y\mathcal{S}_{\mathbf{H}}(z)} \mathrm{d}P_{\mathbf{Y}}(y)\right) \tag{7.3}$$

This theorem is generalized by Girko [79] who derived a fixed-point equation for the Stieltjes transform of large Hermitian matrices $\mathbf{H} = \mathbf{W}\mathbf{W}^{\mathsf{H}}$ when $\mathbf{H}$ has independent entries w_{ij} with variance σ_{ij}^2/N such that the set $\left\{\sigma_{ij}^2\right\}_{i,j}$ is uniformly upper-bounded. In the following, we will extensively use this result to derive the asymptotic network MIMO capacity.

7.3 Fundamental capacity limits

7.3.1 Massively dense network MIMO capacity

We first consider a network MIMO system with n_{T} virtual transmit antennas (one for each base station) and n_{R} virtual receive antennas (one for each mobile terminal). The linear transmission model is

$$\mathbf{y} = \sqrt{\frac{\mathrm{SNR}}{n_{\mathrm{T}}}}\mathbf{H}\mathbf{x} + \mathbf{n} \tag{7.4}$$

with transmit vector of signals sent by the base station $\mathbf{x} \in \mathbb{C}^{n_{\mathrm{T}}}$, receive vector $\mathbf{y} \in \mathbb{C}^{n_{\mathrm{R}}}$ and channel $\mathbf{H} \in \mathcal{M}(\mathbb{C}, n_{\mathrm{R}}, n_{\mathrm{T}})$. The noise vector $\mathbf{n}$ has independent circularly symmetric standard Gaussian entries and SNR is the average Signal-to-Noise-Ratio.

Let the elements of the transmit vector $\mathbf{x}$ be Gaussian with covariance matrix $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}] = \mathbf{\Phi}$. The ergodic achievable rate per sub-channel per cell is given by [69]

$$\mathcal{C}(n_{\mathrm{R}}, n_{\mathrm{T}}) = \mathbb{E}\left[\log\det\left(\mathbb{I}_{n_{\mathrm{R}}} + \frac{\mathrm{SNR}}{n_{\mathrm{T}}}\mathsf{H}\mathbf{\Phi}\mathsf{H}^{\mathsf{H}}\right)\right] \tag{7.5}$$

Following Jakes' model [80], the spatial autocorrelation functions of fading processes h_1 and h_2 experienced by two antennas separated by distance d reads

$$\mathbb{E}[h_1 h_2^*] = J_0(2\pi d/\lambda) \tag{7.6}$$

where $\lambda = c/f_c$ denotes the transmit signal wavelength and $J_0(x)$ is the zero-order Bessel function of the first kind. Thus, by placing several virtual antennas in close proximity their signals tend to be, to some extent, correlated. The correlation function of Jakes' model is depicted in figure 7.1.

In [74], Pollock et al. considered an increasing number of antennas distributed on a uniform circular array of fixed radius. By using bounds on the Bessel function, Pollock derived an approximation of the channel capacity and shows that the capacity bound is independent of $(n_{\mathrm{R}}, n_{\mathrm{T}})$. In the following, we will extend these results using Random Matrix Theory. For a given $\beta \in \mathbb{R}^+$, we will consider that $n_{\mathrm{T}}/n_{\mathrm{R}} \to \beta$ when n_{T} and n_{R} grow large. The entries of $\mathbf{H}$ represent the fading coefficients between each transmit and each receive antenna normalized such that

$$\mathbb{E}\left[\mathrm{tr}\left(\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right] = n_{\mathrm{R}} n_{\mathrm{T}} \tag{7.7}$$

while

$$\mathbb{E}[\|\mathbf{x}\|^2] = n_{\mathrm{T}} \tag{7.8}$$

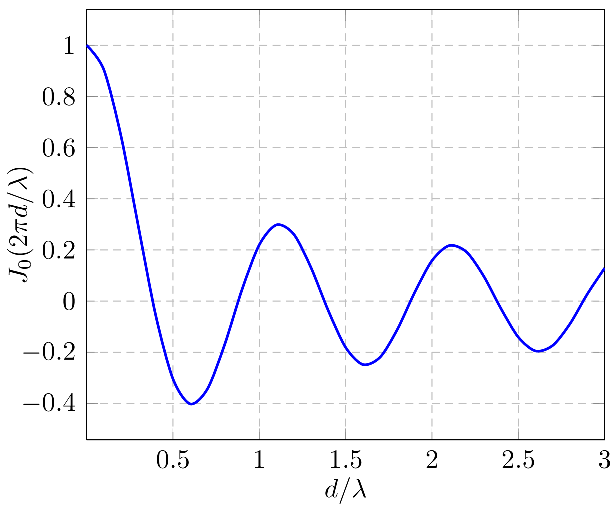


Figure 7.1: Spatial correlation vs. d/λ

It is useful to decompose the input covariance matrix $\mathbf{\Phi} = \mathbb{E}[\mathbf{x}\mathbf{x}^{\mathsf{H}}]$ in its eigenvectors and eigenvalues,

$$\mathbf{\Phi} = \mathbf{V}\mathbf{P}\mathbf{V}^{\mathsf{H}} \tag{7.9}$$

According to the *maximum entropy principle* [81], the most appropriate density function for $\mathbf{H}$, given n_{R}, n_{T}, l and λ, is the classical separable (also termed Kronecker or product) correlation model [82]

$$\mathbf{H} = \mathbf{\Theta}_{\mathbf{R}}^{1/2}\mathbf{H}_w\mathbf{\Theta}_{\mathbf{T}}^{1/2} \tag{7.10}$$

where the deterministic matrices $\mathbf{\Theta}_{\mathbf{T}}$ and $\mathbf{\Theta}_{\mathbf{R}}$ represent the correlation between the virtual transmit antennas and the virtual receive antennas, respectively. The entries of $\mathbf{H}_w$ are i.i.d. standard Gaussian. With statistical channel state information at the transmitter (CSIT), capacity is achieved if the eigenvectors of the input covariance $\mathbf{\Phi}$ coincide with those of $\mathbf{\Theta}_{\mathbf{T}}$. Consequently, denoting $\mathbf{\Lambda}_{\mathbf{T}}$ and $\mathbf{\Lambda}_{\mathbf{R}}$ the diagonal eigenvalue matrices of $\mathbf{\Theta}_{\mathbf{T}}$ and $\mathbf{\Theta}_{\mathbf{R}}$ respectively we have

$$\mathcal{C}(\beta, \mathrm{SNR}) = \lim_{n_{\mathrm{T}}\to\infty} \log\det\left(\mathbb{I} + \frac{\mathrm{SNR}}{n_{\mathrm{T}}}\mathbf{\Lambda}_{\mathbf{R}}^{1/2}\mathbf{H}_w\mathbf{\Lambda}_{\mathbf{T}}^{1/2}\mathbf{P}\mathbf{\Lambda}_{\mathbf{T}}^{1/2}\mathbf{H}_w^{\mathsf{H}}\mathbf{\Lambda}_{\mathbf{R}}^{1/2}\right) \tag{7.11}$$

As a direct consequence of theorem 7.2.1:

Theorem 7.3.1 *[83] The capacity of a Rayleigh-faded channel with separable transmit and receive correlation matrices $\mathbf{\Theta}_{\mathbf{T}}$ and $\mathbf{\Theta}_{\mathbf{R}}$ and statistical CSIT almost surely converges to*

$$\begin{aligned}\frac{\mathcal{C}(\beta, \mathrm{SNR})}{n_{\mathbf{R}}} \to & \beta\mathbb{E}[\log(1 + \mathrm{SNR}\cdot\lambda_{\mathbf{T}}\mathcal{S}(\mathrm{SNR}))] \\ & + \mathbb{E}[\log(1 + \mathrm{SNR}\cdot\lambda_{\mathbf{R}}\Upsilon(\mathrm{SNR}))] \\ & - \beta\cdot\mathrm{SNR}\cdot\mathcal{S}(\mathrm{SNR})\Upsilon(\mathrm{SNR})\log e\end{aligned} \tag{7.12}$$

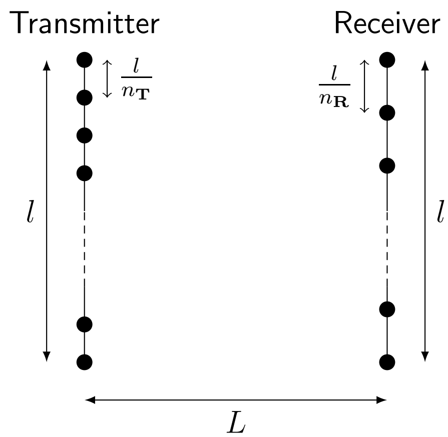


Figure 7.2: One-dimensional antenna array geometry

where

$$\mathcal{S}(\mathrm{SNR}) = \frac{1}{\beta}\mathbb{E}\left[\frac{\lambda_{\mathbf{R}}}{1+\mathrm{SNR}\cdot\lambda_{\mathbf{R}}\Upsilon(\mathrm{SNR})}\right] \tag{7.13}$$

$$\Upsilon(\mathrm{SNR}) = \mathbb{E}\left[\frac{\lambda_{\mathbf{T}}}{1+\mathrm{SNR}\cdot\lambda_{\mathbf{T}}\mathcal{S}(\mathrm{SNR})}\right] \tag{7.14}$$

and the dumb random variables $\lambda_{\mathbf{R}}$, $\lambda_{\mathbf{T}}$ are asymptotically distributed as the diagonal elements of $\mathbf{\Lambda_R}$ and $\mathbf{P\Lambda_T}$ respectively.

7.3.2 Impact of base station geometry and correlation

One-dimensional setup

The antenna setup is depicted in figure 7.2. We consider two uniform linear antenna arrays of length l placed at a distance L. The transmit and receive array is equipped with $n_{\mathbf{T}}$ and $n_{\mathbf{R}}$ antennas, respectively. The correlation matrices $\mathbf{\Theta_T}$ and $\mathbf{\Theta_R}$ have the same form and read

$$\begin{bmatrix} 1 & J_0(\frac{2\pi}{\lambda}\frac{l}{N-1}) & J_0(\frac{2\pi}{\lambda}\frac{2l}{N-1}) & \dots & J_0(\frac{2\pi}{\lambda}\frac{(N-1)l}{N-1}) \\ J_0(\frac{2\pi}{\lambda}\frac{l}{N-1}) & 1 & J_0(\frac{2\pi}{\lambda}\frac{l}{N-1}) & \dots & J_0(\frac{2\pi}{\lambda}\frac{(N-2)l}{N-1}) \\ J_0(\frac{2\pi}{\lambda}\frac{2l}{N-1}) & J_0(\frac{2\pi}{\lambda}\frac{l}{N-1}) & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ J_0(\frac{2\pi}{\lambda}\frac{(N-1)l}{N-1}) & J_0(\frac{2\pi}{\lambda}\frac{(N-2)l}{N-1}) & \dots & \dots & 1 \end{bmatrix} \tag{7.15}$$

with N equal to $n_{\mathbf{T}}$, $n_{\mathbf{R}}$ for $\mathbf{\Theta_T}$, $\mathbf{\Theta_R}$, respectively. The normalized matrices $\frac{1}{n_{\mathbf{R}}}\mathbf{\Theta_R}$ and $\frac{1}{n_{\mathbf{T}}}\mathbf{\Theta_T}$ are *Wiener class* Toeplitz matrices [84], *i.e.*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}|\,[\mathbf{\Theta}]_{1,k}\,| < \infty \tag{7.16}$$

There is no exact expression for the eigenvalues like in the case of a circulant matrix. However, for large $(n_\mathbf{R}, n_\mathbf{T})$ the eigenvalue distribution of a Wiener class Toeplitz matrix converges to that of the circulant matrix, both with the same first row [84]. The set of the eigenvalues of $\frac{1}{n_\mathbf{R}}\mathbf{\Theta_R}$ and $\frac{1}{n_\mathbf{T}}\mathbf{\Theta_T}$ for large $(n_\mathbf{R}, n_\mathbf{T})$ is the image of the function

$$\begin{aligned} F_1 : \ \mathbb{N} \quad & \to \mathbb{R} \\ n \quad & \mapsto \lim_{N\to\infty} \frac{1}{N} \sum_{p=-(N-1)}^{N-1} J_0\left(\frac{2\pi l}{\lambda}\frac{p}{N-1}\right)\cos(2\pi n\frac{p}{N}) \qquad (7.17) \\ & = 2\int_{-1}^{1} J_0\left(\frac{2\pi l}{\lambda}x\right)\cos(2\pi n x)\mathrm{d}x \qquad (7.18) \end{aligned}$$

Since $F(\mathbb{N})$ is a discrete countable set (and not a continuum), the limit eigenvalue distribution of $\mathbf{\Theta_T}$ and $\mathbf{\Theta_R}$ is a sum of Dirac functions

$$p_\nu(\nu) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} \delta\left(\nu - N \cdot F_1(k)\right) \qquad (7.19)$$

At this point it is important to note that the cumulated surface of both virtual antenna arrays must be constant regardless of $n_\mathbf{R}$ and $n_\mathbf{T}$. Hence increasing the number of antennas must lead to a reduction of the individual antenna surface. As a result, the power per receive antenna must scale with $1/n_\mathbf{R}$, hence

$$\mathrm{SNR} = \frac{\mathrm{SNR}'}{n_\mathbf{R}} \qquad (7.20)$$

for a constant total SNR SNR$'$. We first consider the case where no CSIT is available, hence a uniform power allocation over the transmit antennas is optimal (*i.e.* $\mathbf{P} = \mathbf{I}_{n_\mathbf{T}}$). Applying Theorem 7.3.1 and expanding the expectations for large $(n_\mathbf{R}, n_\mathbf{T})$, we have

$$\begin{aligned} \mathcal{C}(\beta, \mathrm{SNR}') =& \beta n_\mathbf{R} \frac{1}{n_\mathbf{T}} \sum_{k=0}^{n_\mathbf{T}} \log(1 + \frac{\rho'}{n_\mathrm{R}} n_\mathrm{T} F_1(k)\mathcal{S}) + n_\mathrm{R}\frac{1}{n_\mathrm{R}}\sum_{k=0}^{n_\mathrm{R}} \log(1+\frac{\rho'}{n_\mathrm{R}} n_\mathrm{R} F_1(k)\Upsilon) \\ & - n_\mathrm{R}\beta \cdot \frac{\mathrm{SNR}'}{n_\mathrm{R}} \cdot \mathcal{S}(\mathrm{SNR}')\Upsilon(\mathrm{SNR}')\log(e) \qquad (7.21) \end{aligned}$$

with

$$\mathcal{S}(\mathrm{SNR}') = \frac{1}{\beta n_\mathbf{R}} \sum_{k=0}^{n_\mathbf{R}} \frac{n_\mathbf{R} F_1(k)}{1 + \mathrm{SNR}' F_1(k)\Upsilon} \quad \text{and} \quad \Upsilon(\mathrm{SNR}') = \frac{1}{n_\mathbf{T}} \sum_{k=0}^{n_\mathbf{T}} \frac{n_\mathbf{T} F_1(k)}{1 + \mathrm{SNR}'\beta F_1(k)\mathcal{S}} \qquad (7.22)$$

In the limit, this becomes

$$\begin{aligned} \mathcal{C}(\beta, \mathrm{SNR}') \to & \sum_{k=0}^{+\infty} \log(1 + \mathrm{SNR}'\beta F_1(k)\mathcal{S}) + \sum_{k=0}^{+\infty} \log(1 + \mathrm{SNR}' F_1(k)\Upsilon) \\ & - \beta \mathrm{SNR}' \mathcal{S}(\mathrm{SNR}')\Upsilon(\mathrm{SNR}')\log(e) \qquad (7.23) \end{aligned}$$

with

$$\mathcal{S}(\mathrm{SNR}') = \frac{1}{\beta}\sum_{k=0}^{+\infty}\frac{F_1(k)}{1+\mathrm{SNR}'F_1(k)\Upsilon} \tag{7.24}$$

$$\Upsilon(\mathrm{SNR}') = \sum_{k=0}^{+\infty}\frac{F_1(k)}{1+\mathrm{SNR}'\beta F_1(k)\mathcal{S}} \tag{7.25}$$

where $\forall k \in \mathbb{N},\ F_1(k) \geq 0$ as they are eigenvalues of a covariance matrix. Also $\sum_{k=0}^{+\infty} F_1(k) = \frac{1}{n_{\mathbf{R}}}\mathrm{tr}(\mathbf{\Theta_R}) = \mathbf{1}$. This implies that $\mathcal{S}$ and Υ are finite and therefore the total capacity $\mathcal{C}$ is also finite.

Further note that (7.19) only depends on the system parameters through the ratio l/λ. This leads to the conclusion that the network MIMO capacity limit depends only on the ratio l/λ and β when CSIT is absent.

Consider now the case of perfect CSIT. Here, it is optimal to distribute the power according to the water-filling solution [85]. That is, only *sufficiently strong* eigenmodes of the channel (7.10) are used for transmission. If we allocate the power constrained by (7.8) on the dominating channel eigenmodes (*i.e.* the relevant eigenvalues of $\frac{1}{n_{\mathbf{T}}}\mathbf{\Theta}_{\mathbf{R}}^{1/2}\mathbf{H}_w\mathbf{\Theta_T}\mathbf{H}_w^{\mathsf{H}}\mathbf{\Theta}_{\mathbf{R}}^{1/2}$), then for large $n_{\mathbf{T}}$, the capacity grows unbounded. As a result, increasing the number of antennas at either side of the transmission allows to achieve *arbitrarily high* capacity under the assumption of perfect CSIT.

Two-dimensional setup

The previous scheme can be extended to a surface area. We now increase the density of virtual antennas uniformly along each dimension of the surface. Consider a rectangular surface of respective height and width l_x and l_y. Then, equation (7.19) has an equivalent version in two dimensions,

$$p_\nu(\nu) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N}\delta(\nu - N\cdot F_2(k)) \tag{7.26}$$

with

$$\begin{aligned} F_2(k) =& 2\int_{-l_x}^{l_x}\int_{-l_y}^{l_y} J_0\left(\frac{2\pi l_x}{\lambda}\sqrt{u_x^2+\left(\frac{l_y}{l_x}\right)^2 u_y^2}\right) \\ & \times \cos(2\pi\nu(u_x+u_y))\mathrm{d}u_x\mathrm{d}u_y \end{aligned} \tag{7.27}$$

From (7.26), one verifies that the final capacity formulation depends on the two constant values l_x/λ and l_x/l_y, or similarly on the two ratios l_x/λ and l_y/λ. Note that the capacity for a given surface might then differ depending on the shape of the surface.

7.3.3 MIMO-GBC capacity

We assume a GBC generated by a base station with multiple antennas and many non-cooperative single-antenna receivers. It has been shown in [86] that the capacity region is achieved by dirty-paper coding (DPC). However, to derive closed-form expressions, we restrict our analysis in the following to suboptimal linear precoding techniques. As the receivers are uncorrelated, the GBC channel model is

$$\mathbf{H} = \mathbf{H}_w \mathbf{\Theta}_\mathbf{T}^{1/2} \tag{7.28}$$

and the transmitted signal $\mathbf{x}$ is obtained by

$$\mathbf{x} = \mathbf{G}\mathbf{u} \tag{7.29}$$

where the symbol vector $\mathbf{u} \in \mathbb{C}^{n_\mathbf{R}}$ has unit power, and $\mathbf{G} \in \mathcal{M}(\mathbb{C}, n_\mathbf{T}, n_\mathbf{R})$.

ZF-beamforming

Zero-Forcing (ZF) beamforming is a mere channel inversion precoding. With the same notation as in previous sections, the channel model reads

$$\mathbf{y} = \sqrt{\frac{\mathrm{SNR}}{n_\mathbf{T}}} \mathbf{H}\mathbf{x} + \mathbf{n} \tag{7.30}$$

with

$$\mathbf{x} = \begin{cases} \alpha \mathbf{H}'^{-1}\mathbf{u} & , \text{ if } \beta = 1 \\ \alpha \left(\mathbf{H}'^{\mathsf{H}}\mathbf{H}'\right)^{-1} \mathbf{H}'^{\mathsf{H}}\mathbf{u} & , \text{ if } \beta > 1 \end{cases} \tag{7.31}$$

for $\mathbf{H}' = \sqrt{\frac{1}{n_\mathbf{T}}}\mathbf{H}$. The parameter α is set to fulfill the transmission power constraint (7.8) which leads to

$$\alpha^2 = \frac{1}{\frac{1}{n_\mathrm{T}} \mathrm{tr} \left(\mathbf{H}'\mathbf{H}'^{\mathsf{H}}\right)^{-1}} \tag{7.32}$$

where

$$\frac{1}{n_\mathbf{T}} \mathrm{tr} \left(\mathbf{H}'\mathbf{H}'^{\mathsf{H}}\right)^{-1} \to \int \frac{1}{\nu} f(\nu) \mathrm{d}\nu \tag{7.33}$$

with f the empirical distribution of $\mathbf{H}'\mathbf{H}'^{\mathsf{H}}$. We recognize in (7.33) the Stieltjes transform of $f(x)$ in $x = 0$.

To the authors' knowledge, in contrast to [87] where no power limitation is imposed on $\mathbf{x}$, no asymptotic expression for α is known when $(n_\mathbf{R}, n_\mathbf{T})$ grow large.

Recall that $\mathbf{H}\mathbf{H}^{\mathsf{H}} = \mathbf{H}_w \mathbf{\Theta}_\mathbf{T} \mathbf{H}_w^{\mathsf{H}}$. Thus by diagonalizing $\mathbf{\Theta}_\mathbf{T} = \mathbf{V}\mathbf{\Lambda}_\mathbf{T}\mathbf{V}^{\mathsf{H}}$ with unitary matrix $\mathbf{V}$, we have

$$\mathbf{H}'\mathbf{H}'^{\mathsf{H}} = \left(\frac{1}{\sqrt{n_\mathrm{T}}}\mathbf{H}_w\mathbf{V}\right) \mathbf{\Lambda}_\mathbf{T} \left(\frac{1}{\sqrt{n_\mathrm{T}}}\mathbf{V}^{\mathsf{H}}\mathbf{H}_w^{\mathsf{H}}\right) \tag{7.34}$$

where the entries of $\frac{1}{\sqrt{n_\mathbf{T}}}\mathbf{H}_w\mathbf{V}$ are i.i.d. with zero mean and variance $\frac{1}{n_\mathrm{T}}$, and $\mathbf{\Lambda_T}$ is distributed as in (7.19). Applying theorem 7.2.1, we prove the existence of $\mathcal{S}_{\mathbf{H}'\mathbf{H}'^\mathsf{H}}$, when $n_\mathrm{T}/n_\mathrm{R} \to \beta$, that satisfies

$$\mathcal{S}_{\mathbf{H}'\mathbf{H}'^\mathsf{H}}(z) = \left(-z + \beta \int \frac{\nu p_\nu(\nu)}{1+\nu\cdot\mathcal{S}_{\mathbf{H}'\mathbf{H}'^\mathsf{H}}(z)}\mathrm{d}\nu\right)^{-1} \tag{7.35}$$

Expanding $\mathbf{x}$ according to (7.4), one obtains parallel non-interfering channels with the per-user capacity

$$\begin{aligned}\mathcal{C}_u(\beta, \mathrm{SNR}) &= \log(1 + \mathrm{SNR}\alpha^2) &(7.36)\\ &= \log(1 + \mathrm{SNR}\mathcal{S}_{\mathbf{H}'\mathbf{H}'^\mathsf{H}}(0)^{-1}) &(7.37)\end{aligned}$$

In our specific correlation scenario, this capacity limit is in fact zero. Indeed, if $n_\mathbf{T} = n_\mathbf{R}$ we have

$$\frac{1}{n_\mathbf{T}}\mathrm{tr}\left(\mathbf{H}'\mathbf{H}'^\mathsf{H}\right)^{-1} = \frac{1}{n_\mathbf{T}}\mathrm{tr}\left((\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w)^{-1}\mathbf{\Lambda}_\mathbf{T}^{-1}\right) \tag{7.38}$$

where $\tilde{\mathbf{H}}_w = \mathbf{H}_w\mathbf{V}$ is a Gaussian random matrix with entries of variance $1/n_\mathbf{T}$.

Lemma 7.3.1 *For any two Hermitian $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ with eigenvalues $\lambda_i(\mathbf{A})$ and $\lambda_i(\mathbf{B})$ respectively arranged in decreasing order,*

$$\mathrm{tr}(\mathbf{AB}) \geq \sum_{i=1}^{n} \lambda_i(\mathbf{A})\lambda_{n-i+1}(\mathbf{B}) \tag{7.39}$$

From lemma 7.3.1, we have

$$\mathrm{tr}\left((\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w)^{-1}\mathbf{\Lambda}_\mathbf{T}^{-1}\right) \geq \sum_{i=1}^{n_\mathbf{R}} \lambda_i((\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w)^{-1})\lambda_{n-i+1}(\mathbf{\Lambda}_\mathbf{T}^{-1}) \tag{7.40}$$

The eigenvalues of $\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w$ are known [76] to be asymptotically distributed as the *Marčhenko-Pastur* law on a bounded (positive) support excluding zero. Therefore the eigenvalues of $(\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w)^{-1}$ are also bounded on a finite positive support. Denote λ_{min} the minimum of those eigenvalues, we have

$$\mathrm{tr}\left((\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w)^{-1}\mathbf{\Lambda}_\mathbf{T}^{-1}\right) \geq \lambda_{min}\sum_{i=1}^{n_\mathbf{R}} \lambda_i(\mathbf{\Lambda}_\mathbf{T}^{-1}) \tag{7.41}$$

Observing that

$$\lambda_{n_\mathbf{R}}(\mathbf{\Lambda_T}) = \sum_{p=-(n_\mathbf{R}-1)}^{n_\mathbf{R}-1} J_0\left(\frac{2\pi l}{\lambda}\frac{p}{n_\mathbf{R}-1}\right)\cos(2\pi\frac{p}{N}) \to 0 \tag{7.42}$$

we conclude

$$\mathrm{tr}\left((\tilde{\mathbf{H}}_w^\mathsf{H}\tilde{\mathbf{H}}_w)^{-1}\mathbf{\Lambda}_\mathbf{T}^{-1}\right) \to +\infty \tag{7.43}$$

Therefore, $\alpha^2 \to 0$ and the ZF capacity goes to zero for increasing $(n_\mathbf{R}, n_\mathbf{T})$ and $\beta = 1$. The case $n_\mathbf{T} > n_\mathbf{R}$ can be solved by dividing $\tilde{\mathbf{H}}_w$ in a two blocks of size $n_\mathbf{R}\times n_\mathbf{R}$ and $(n_\mathbf{T}-n_\mathbf{R})\times n_\mathbf{R}$ where the capacity limit for the former already grows to infinity.

MMSE-beamforming

Let us consider regularized ZF-beamforming. The system model in (7.30) becomes

$$\mathbf{x} = \left(\mathbf{H'}^{\mathsf{H}}\mathbf{H'} + \alpha\mathbf{I}_{n_\mathbf{T}}\right)^{-1}\mathbf{H'}^{\mathsf{H}}\mathbf{u} \tag{7.44}$$

When $\alpha = 0$, we fall back the ZF solution. The parameter α is set so to fulfill the transmission power constraint (7.8) which leads to

$$\begin{aligned}
1 &= \frac{1}{n_\mathbf{T}}\mathrm{tr}\left(\left(\mathbf{H'}^{\mathsf{H}}\mathbf{H'} + \alpha\mathbf{I}\right)^{-1}\mathbf{H'}^{\mathsf{H}}\mathbf{H'}\left(\mathbf{H'}^{\mathsf{H}}\mathbf{H'} + \alpha\mathbf{I}\right)^{-1}\right) && (7.45)\\
&= \frac{1}{n_\mathbf{T}}\mathrm{tr}\left(\left(\mathbf{H'}^{\mathsf{H}}\mathbf{H'} + \alpha\mathbf{I}\right)^{-2}\mathbf{H'}^{\mathsf{H}}\mathbf{H'}\right) && (7.46)\\
&\to \int \frac{\nu}{(\nu+\alpha)^2} f(\nu)\mathrm{d}\nu && (7.47)\\
&= \int \left(\frac{1}{(\nu+\alpha)} - \frac{\alpha}{(\nu+\alpha)^2}\right) f(\nu)\mathrm{d}\nu && (7.48)\\
&= \mathcal{S}_{\mathbf{H'}\mathbf{H'}^{\mathsf{H}}}(-\alpha) + \alpha\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{S}_{\mathbf{H'}\mathbf{H'}^{\mathsf{H}}}(-\alpha) && (7.49)
\end{aligned}$$

The received signal can be written as

$$\mathbf{y} = \sqrt{\rho}\cdot\mathbf{H'}\left(\mathbf{H'}^{\mathsf{H}}\mathbf{H} + \alpha\mathbf{I}\right)^{-1}\mathbf{H'}^{\mathsf{H}}\mathbf{u} + \mathbf{n} \tag{7.50}$$

Let us denote $\mathbf{H'}^{\mathsf{H}} = [\mathbf{h}_1, \ldots, \mathbf{h}_{n_\mathbf{R}}]$. We will focus on user i without loss of generality. The received symbol of user i is

$$\begin{aligned}
y_i &= \sqrt{\rho}\cdot\mathbf{h}_i^{\mathsf{H}}\left(\mathbf{H'}^{\mathsf{H}}\mathbf{H'} + \alpha\mathbf{I}\right)^{-1}\mathbf{h}_i u_i\\
&+ \sum_{k=1,k\neq i}^{n_\mathbf{R}} \mathbf{h}_i^{\mathsf{H}}\left(\mathbf{H'}^{\mathsf{H}}\mathbf{H'} + \alpha\mathbf{I}\right)^{-1}\mathbf{h}_k u_k\\
&+ \mathbf{n} && (7.51)
\end{aligned}$$

Lemma 7.3.2 *[88] Let* $\mathbf{A}$ *be a deterministic* $N \times N$ *complex matrix with uniformly bounded spectral radius for all* N*. Let* $\mathbf{x} = \frac{1}{\sqrt{N}}[x_1, \ldots, x_N]^T$ *where the* $\{x_i\}$ *are i.i.d complex random variables with zero mean, unit variance and finite eighth moment. Then*

$$\mathbb{E}\left[\mid \mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{tr}\mathbf{A} \mid^4\right] \leq \frac{c}{N^2} \tag{7.52}$$

where c *is a constant that does not depend on* N *or* $\mathbf{A}$*.*

Corollary 7.3.1 *This result ensures that*

$$\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{tr}\mathbf{A} \to 0 \tag{7.53}$$

almost surely.

Henceforth we write $\mathbf{U}_i^{\mathsf{H}} = [\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}, \mathbf{h}_{i+1}, \ldots, \mathbf{h}_{n_\mathbf{R}}]$ (in other words, we remove column i). Applying the matrix inversion lemma yields

$$\mathbf{h}_i^{\mathsf{H}} \left(\mathbf{H}'^{\mathsf{H}}\mathbf{H}' + \alpha\mathbf{I}\right)^{-1} = \frac{\mathbf{h}_i^{\mathsf{H}} \left(\mathbf{U}_i^{\mathsf{H}}\mathbf{U}_i + \alpha\mathbf{I}\right)^{-1}}{1 + \mathbf{h}_i^{\mathsf{H}} \left(\mathbf{U}_i^{\mathsf{H}}\mathbf{U}_i + \alpha\mathbf{I}\right)^{-1}\mathbf{h}_i}$$

As the elements of $\mathbf{h}_i$ are i.i.d. (due to the one sided correlation assumption), we can use lemma 7.3.2

$$\mathbf{h}_i^{\mathsf{H}} \left(\mathbf{H}'^{\mathsf{H}}\mathbf{H}' + \alpha\mathbf{I}\right)^{-1}\mathbf{h}_i \to \frac{1}{n_\mathbf{T}}\mathrm{tr}\left(\mathbf{H}'^{\mathsf{H}}\mathbf{H}'\right) \tag{7.54}$$

Asymptotically, the removal of a single column in the large matrix $\mathbf{H}'$ does not affect $\mathrm{tr}(\mathbf{H}'^{\mathsf{H}}\mathbf{H}')$, we have

$$\mathbf{h}_i^{\mathsf{H}} \left(\mathbf{U}_i^{\mathsf{H}}\mathbf{U}_i + \alpha\mathbf{I}\right)^{-1}\mathbf{h}_i \to \mathbf{h}_i^{\mathsf{H}} \left(\mathbf{H}'^{\mathsf{H}}\mathbf{H}' + \alpha\mathbf{I}\right)^{-1}\mathbf{h}_i \tag{7.55}$$

hence

$$\mathbf{h}_i^{\mathsf{H}} \left(\mathbf{H}'^{\mathsf{H}}\mathbf{H}' + \alpha\mathbf{I}\right)^{-1} \to \frac{\mathbf{h}_i^{\mathsf{H}} \left(\mathbf{U}_i^{\mathsf{H}}\mathbf{U}_i + \alpha\mathbf{I}\right)^{-1}}{1 + \mathcal{S}_{\mathbf{H}'\mathbf{H}'^{\mathsf{H}}}(-\alpha)}$$

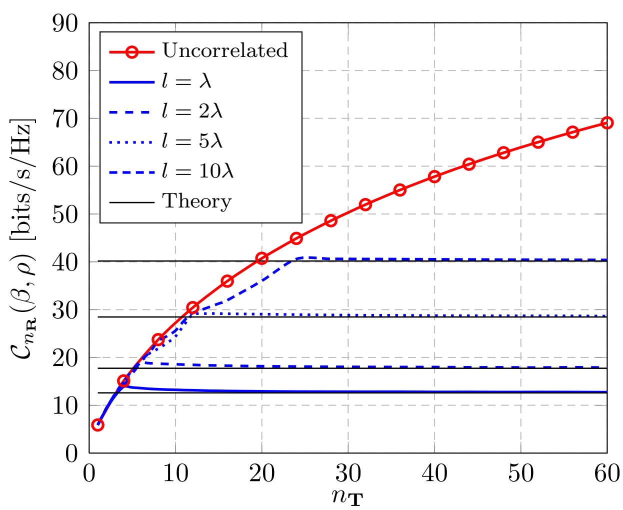


Figure 7.3: Ergodic network MIMO capacity $\mathcal{C}_{n_\mathbf{R}}(\beta, \rho)$ *without* CSIT for different l/λ, $n_\mathbf{R} = n_\mathbf{T}$, $\mathrm{SNR}' = 20$ dB

Denote $\gamma = (1 + \mathcal{S}_{\mathbf{H}\mathbf{H}^{\mathsf{H}}}(-\alpha))^{-2}$. The expression for the SINR is therefore given by

$$\mathsf{SINR}_i = \frac{\rho\gamma \mathbf{h}_i^{\mathsf{H}} \mathbf{W}_i^2 \mathbf{h}_i}{\rho\gamma \mathbf{h}_i^{\mathsf{H}} \mathbf{W}_i \mathbf{U}_i^{\mathsf{H}} \mathbf{U}_i \mathbf{W}_i \mathbf{h}_i + 1} \tag{7.56}$$

with $\mathbf{W}_i = \left(\mathbf{U}_i^{\mathsf{H}} \mathbf{U}_i + \alpha \mathbf{I}\right)^{-1}$. In the limit this leads to a *user-independent* SINR

$$\mathsf{SINR} \to \frac{\rho\gamma \mathcal{S}^2_{\mathbf{H}'\mathbf{H}'^{\mathsf{H}}}(-\alpha)}{\rho\gamma \left(\mathcal{S}_{\mathbf{H}'\mathbf{H}'^{\mathsf{H}}}(-\alpha) + \alpha \frac{\mathrm{d}}{\mathrm{d}x} \mathcal{S}_{\mathbf{H}'\mathbf{H}'^{\mathsf{H}}}(-\alpha)\right) + 1} \tag{7.57}$$

The corresponding per-user capacity is

$$\mathcal{C}_u(\beta, \mathrm{SNR}) = \log(1 + \mathsf{SINR}) \tag{7.58}$$

Diagonalizing $\mathbf{U}^{\mathsf{H}}\mathbf{U}$, we observe that the numerator in (7.56) converges to finite strictly positive values (for the regularization term α ensures that no term diverges). However, as already noted, the strongest eigenvalue of $\mathbf{\Theta}_{\mathbf{T}}$ grows linearly with $n_{\mathbf{T}}$, hence, with to lemma 7.3.1, the denominator grows to infinite for large $n_{\mathbf{T}}$. This proves that the per-user capacity goes to zero.

Hence, for large $(n_{\mathbf{R}}, n_{\mathbf{T}})$ the MMSE-beamforming algorithm yields zero per-user capacity. Therefore both MMSE beamforming and ZF beamforming achieve asymptotically zero per-user capacity.

As a consequence, it turns out that additional antennas might impair the achievable transmission rate. This is explained by the fact that loading power on more and more correlated antennas, instead of available channel modes, is an inefficient power allocation strategy.

7.4 Network simulations and results

Let us first consider the network MIMO scenario with dense antenna arrays at both transmitter and receiver side. Figures 7.3 and 7.4 present the results of ergodic capacities found by numerical simulation.

$$\mathcal{C}_{n_{\mathbf{R}}}(\beta, \rho) = n_{\mathbf{R}} \mathbb{E}\left[\log\det\left(\mathbf{I} + \frac{\mathrm{SNR}}{n_{\mathbf{T}}} \mathbf{\Lambda}_{\mathbf{R}}^{1/2} \mathbf{H}_w \mathbf{\Lambda}_{\mathrm{T}}^{1/2} \mathbf{P} \mathbf{\Lambda}_{\mathrm{T}}^{1/2} \mathbf{H}_w^{\mathsf{H}} \mathbf{\Lambda}_{\mathbf{R}}^{1/2}\right)\right]$$

In figure 7.3, we allocate equal power (*i.e.* $\mathbf{P} = \mathbf{I}_{n_{\mathbf{T}}}$) to the transmit symbols. We observe, as previously concluded, that the capacity saturates for large $(n_{\mathbf{T}}, n_{\mathbf{R}})$. In addition we provide the theoretical limits derived from (7.11) (which are obtained by solving numerically (7.24),(7.25)). Note that the capacity increases first to a maximum for small $(n_{\mathbf{R}}, n_{\mathbf{T}})$ and then decreases to the capacity limit. In figure 7.4, we apply water-filling (*i.e.* loading the transmit power on the dominant eigenmodes of the channel), which leads to a non-saturating capacity.

Let us now consider the MIMO-GBC with *uncorrelated* transmitters/receivers and ZF-beamforming. As has been shown in [87] the sum capacity is saturating if $\beta = 1$ and growing

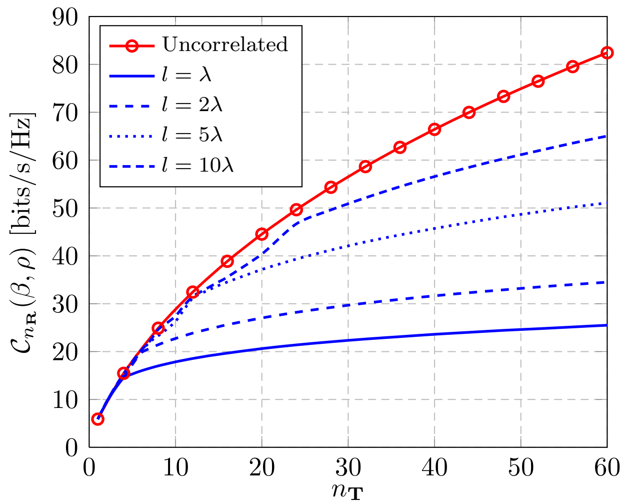


Figure 7.4: Ergodic Network MIMO capacity $\mathcal{C}_{n_\mathbf{R}}(\beta, \rho)$ *with perfect* CSIT for different l/λ, $n_\mathbf{R} = n_\mathbf{T}$, $\mathrm{SNR}' = 20$ dB

linearly with $n_\mathbf{R}$ when $\beta > 1$ which is in accordance with figure 7.6. From this figure we further observe that the sum capacity is going to zero in case of correlation between the transmit antennas. Figure 7.5 shows the corresponding per-user capacity.

In figures 7.7 and 7.8 we apply MMSE-beamforming. Since no closed-form solution for α under the constraint (7.49) is available, the optimal α is found by exhaustive search. We observe again that the per-user capacity is going asymptotically to zero, which is in accordance with equations (7.57) and (7.58). The same observation can be made for the sum capacity in figure 7.8. Both, the per-user capacity and the sum capacity are decreasing less rapidly for large $(n_\mathbf{R}, n_\mathbf{T})$ than in the case of ZF-beamforming.

7.5 Discussion

A few limitations are worth mentioning about our previous conclusions. In the MIMO case we stated that, with perfect CSIT, the channel capacity grows unbounded even with a strong virtual antenna correlation at the transmitter side. This might indicate that densifying the array of virtual transmit antennas is the preferred option to increase the capacity (rather than increasing the transmitted power or the channel bandwidth). However, perfect CSIT implies that the receiver has to feed back channel information to the transmitter (either as pilot sequences or as directly quantized CSI). For a dense network MIMO system, this introduces an enormous feedback overhead and is thus reducing the achievable throughput.

The same conclusion holds for channel state information at the receiver (CSIR). As Tse demonstrated [89], the capacity with perfect CSIR is limited by the coherence time of the channel. If the number of virtual antennas grows, one needs to estimate more and more degrees of freedom with less power. An optimal trade-off must then be found between increasing the number of antennas (and thus the capacity) and decreasing the amount of channel state information required for reliable transmission. However, if the channel coher-

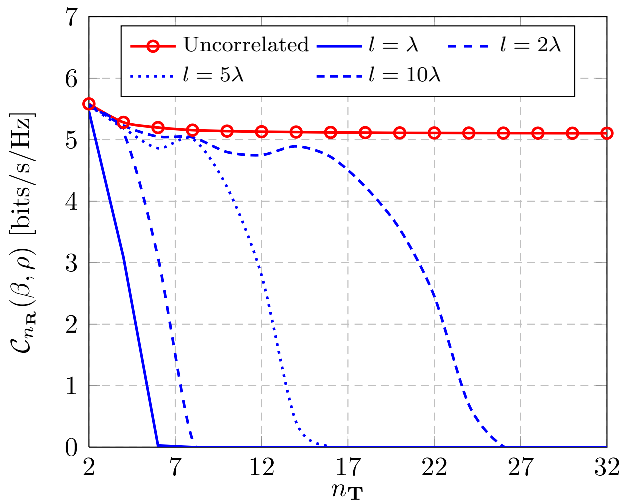


Figure 7.5: Ergodic MIMO-GBC-ZF per-user capacity $\mathcal{C}_{n_\mathbf{R}}(\beta, \rho)$ for different l/λ, $n_\mathbf{T} = \frac{3}{2}n_\mathbf{R}$

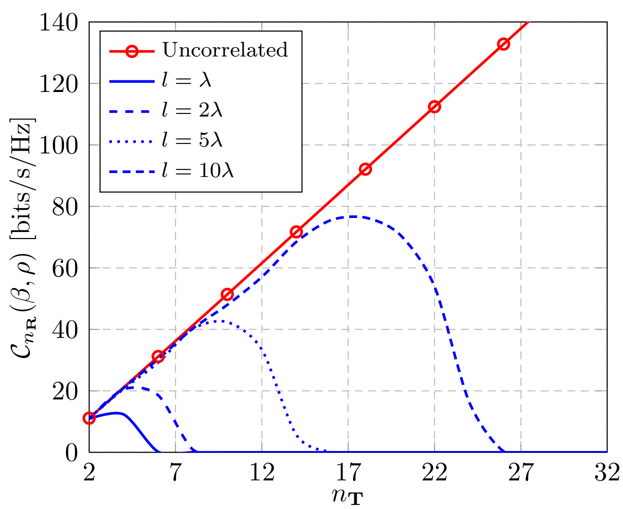


Figure 7.6: Ergodic MIMO-GBC-ZF sum capacity $n_\mathbf{R}\mathcal{C}_{n_\mathbf{R}}(\beta, \rho)$ for different l/λ, $n_\mathbf{T} = \frac{3}{2}n_\mathbf{R}$

ence time is infinite and a long synchronization stage prior to data transmission is allowed, then the channel capacity can effectively go unbounded. The only limitation that would appear lies in the physical ability to design a dense array of virtual antennas on a limited surface. In addition, a dense scattering environment is necessary to assure that the assumed channel model is accurate.

7.6 Conclusions

In this work we analyzed the asymptotic capacity of the dense multiple antenna configurations. For the network MIMO channel we have shown that in the absence of CSIT, the capacity is bounded and related to the ratio between the size of the antenna array and the transmit signal wavelength. The capacity grows unbounded if perfect CSIT is available. In

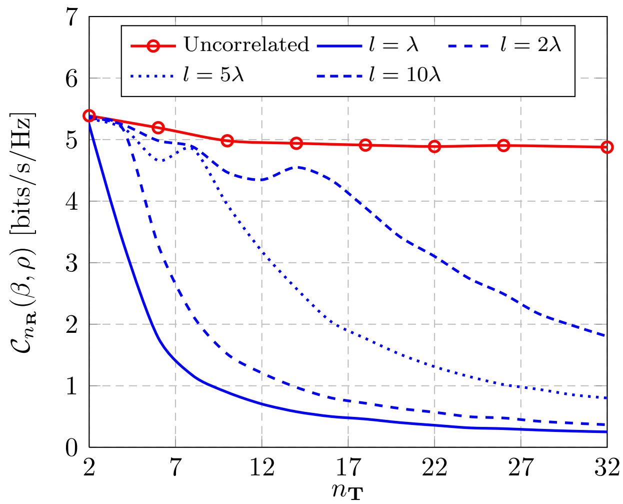


Figure 7.7: Ergodic MIMO-GBC-MMSE per-user capacity $\mathcal{C}_{n_\mathbf{R}}(\beta, \rho)$ for different l/λ, $n_\mathbf{T} = \frac{3}{2}n_\mathbf{R}$

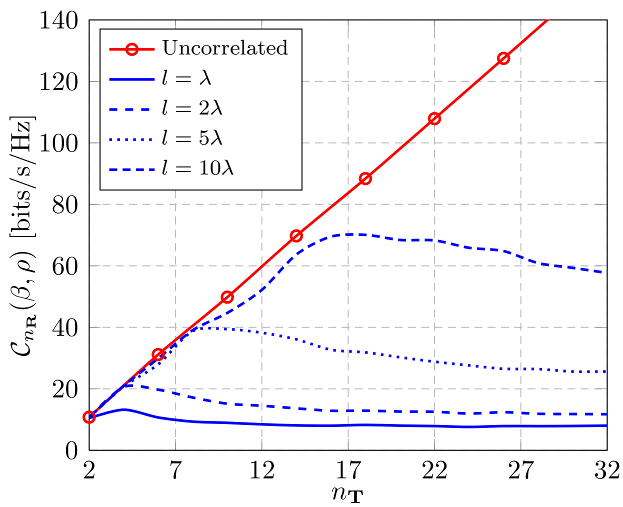


Figure 7.8: Ergodic MIMO-GBC-MMSE sum capacity $n_\mathbf{R}\mathcal{C}_{n_\mathbf{R}}(\beta, \rho)$ for different l/λ, $n_\mathbf{T} = \frac{3}{2}n_\mathbf{R}$

case of the dense MIMO broadcast channel the per-user capacity goes asymptotically to zero for ZF-beamforming as well as for MMSE-beamforming.

Chapter 8

Mobile Association and Optimal Placement of Base Stations

8.1 Introduction

We consider the case where intelligent mobile terminals capable of accessing multiple radio access technologies will decide for themselves the wireless access technology to use and the access point to which to connect. We consider that these capabilities should be taken into account in the design and strategic planning of wireless networks. We also consider the global optimization problem to minimize the total power of the network in the downlink and in the uplink context.

We propose a new framework for mobile association problems using optimal transport theory, a theory that has prove to be useful on many economical context [90], [91], [92], as well as in the road traffic community [93]. There is a number of papers on "optimal transport" (see [94], and reference therein) however the authors in [94] consider an optimal selection of routes but do not use the rich theory of optimal transport. To the best of the authors knowledge optimal transport theory has never been used in the telecommunication community.

The remaining of this chapter is organized as follows. Section 8.2 presents the formulation of the problem of minimizing the power under quality of service constraint from different perspectives. In Section 8.3 we give some basics in optimal transport theory. We then address the problem

- for the downlink case where we considered two different policies: round robin scheduling policy (also known as time fair allocation policy) and rate fair allocation policy which are defined in section 8.2 and studied precisely in sections 8.4 and 8.6 as well as the fairness problem (detailed in section 8.5) with uniform and non-homogeneous distribution of users, and

- for uplink case where we study the optimal cell association with uniform and non-homogeneous distribution of users.

In Section 8.8 we give numerical examples in both one dimensional and two-dimensional mobile distribution. Section 8.9 concludes the paper.

8.2 The model

Consider a grid area network D with large number of mobile terminals distributed with a square integrable distribution of $\lambda(x,y)$ scaled so that $\iint_{\mathrm{D}} \lambda(x,y)\,dx\,dy = 1$. Then the number of users in an area A will be $N\left(\iint_A \lambda(x,y)\right)$ where N is the total number of mobile terminals.

Examples of the distribution of users $\lambda(x,y)$:

1. If the users are distributed uniformly in the network, then $\lambda(x,y) = 1/\bar{D}$ where $\bar{D}$ is the total area of the network.

2. If the users are distributed according to different levels of population density, then

$$\lambda(x,y) = \begin{cases} \lambda_{\mathrm{HD}} \text{ if } (x,y) \text{ is at a High Density region,} \\ \lambda_{\mathrm{ND}} \text{ if } (x,y) \text{ is at a Normal Density region,} \\ \lambda_{\mathrm{LD}} \text{ if } (x,y) \text{ is at a Low Density region.} \end{cases}$$

   where λ_{HD}, λ_{ND}, and λ_{HD} are defined similarly to 1).

3. If the distribution of the users is radial with more mobile terminals in the center and less mobile terminals in the suburban areas then $\lambda(x,y) = \frac{R_D^2-(x^2+y^2)}{K_D}$, where R_D is the radius of the network and K_D is a coefficient of normalization.

4. If the distribution of users is a Poisson process with intensity ν, then

$$\lambda(x,y) = e^{-\nu\pi r^2}$$

   where r is the polar coordinate representation of (x,y). This particular case has been examined in [95].

Notice that the distribution of users $\lambda(x,y)$ considered in our work is more general than all the examples mentioned above.

We assume that in this grid area network there are K base stations $\mathrm{BS}_1, \mathrm{BS}_2, \ldots, \mathrm{BS}_K$ located at positions $(x_1,y_1),(x_2,y_2)\ldots,(x_K,y_K)$. For the uplink case (transmission from mobile terminals to base stations) we consider the SINR (Signal to Interference plus Noise Ratio). However, we assume for the downlink case (transmission from base stations to mobile terminals) that between neighboring base stations, they transmit in orthogonal channels (such as in OFDMA), and the interference between base stations that are far from each other is negligible, so instead of considering the SINR (Signal to Interference plus Noise Ratio) we consider the SNR (Signal to Noise Ratio).

Our objective is to determine the optimal mobile association to each base station in order to minimize the total power of the network needed to maintain an average throughput of $\bar{\theta}(x,y) > 0$ for each mobile of the network located at position (x,y). We also determine the equilibrium situation where the mobile terminals decide for themselves with which base station to connect in order to maximize their rate.

8.2.1 Downlink case

Consider in the downlink case that when the base station BS_i transmits to a mobile terminal located at position (x,y), it uses power $P_i(x,y)$. Each base station BS_i is going to transmit to the mobiles distributed within its cell C_i (the mobile terminals associated to BS_i) to be determined.

Denote by N_i the quantity of mobiles that are assigned to base station BS_i. If the quantity of mobiles is greater than some M (for example, the number of possible carriers in WiMAX is around 2048, so in this case $M = 2048$) then we consider a penalization cost function given by

$$\begin{cases} 0 & \text{if } N_i \leq M, \\ h(N_i - M) & \text{if } N_i > M. \end{cases}$$

We will assume that h is a non-decreasing and convex function. We analyze the case $N_i \leq M$ but for the resolution in section 8.4 we will remove this assumption. As each cell C_i of the network contain a large number of mobiles continuously distributed with a distribution of $\lambda(x,y)$ then the quantity of mobiles assigned to base station BS_i will be

$$N_i = N \iint_{C_i} \lambda(x,y)\, dx\, dy. \tag{8.1}$$

Notice that $\sum_{i=1}^{K} N_i = N$ so each mobile terminal is associated to one base station in the network. The power received at a mobile terminal located at position (x,y) from base station BS_i is given by $P_i(x,y)h_i(x,y)$ where $h_i(x,y)$ is the channel gain. We shall further assume that it corresponds to the path loss given by

$$h_i(x,y) = (R^2 + d_i(x,y)^2)^{-\mathcal{P}/2} \tag{8.2}$$

where $\mathcal{P}$ is the path loss exponent [96], R is the high of the base station, and $d_i(x,y)$ is the Euclidean distance between a mobile located at position (x,y) and the base station BS_i located at (x_i, y_i), *i.e.*, $d_i(x,y) = \sqrt{(x_i - x)^2 + (y_i - y)^2}$.

The SNR received at mobile terminals located at position (x,y) in cell C_i to be determined is given by

$$\mathrm{SNR}_i(x,y) = \frac{P_i(x,y)h_i(x,y)}{\sigma^2}, \tag{8.3}$$

where σ^2 is the noise power.

We assume that the instantaneous mobile throughput is given by the following expression, which is based on Shannon's capacity theorem[97]:

$$\theta_i(x,y) = \log(1 + \mathrm{SNR}_i(x,y)).$$

Suppose that we want to satisfy an average throughput for mobile terminals located at position (x, y) given by $\bar{\theta}(x, y) > 0$.

We shall consider two different policies:

1. the policy that each base station BS_i devotes an equal fraction of time for transmission to each of its mobile terminals located within its cell C_i. We denote this policy as *round robin scheduling policy*.

2. the policy where each base station BS_i will maintain a constant power P_i sent to the mobile terminals within its cell. However, each base station will modify the fraction of time allowed to mobile terminals with different channel gains, in order that the average SNR of $\Theta(x, y)$ is satisfied for each mobile located at position (x, y). We denote this policy as *rate fair allocation policy*

For more information about this type of policies in the one dimensional case see [98].

Round robin scheduling policy

- Global Optimization

Following this policy each base station BS_i devotes an equal fraction of time for transmission to each mobile terminal located within its cell C_i. From equation (8.1) we have that the number of mobiles located in cell C_i is $N_i(C_i)$. As we are dividing our time of service proportional to the quantity of users N_i inside cell C_i then the throughput following the round robin scheduling policy will be given by:

$$\theta_i^{\mathrm{RR}}(x, y) = \frac{1}{N_i} \log(1 + \mathrm{SNR}_i(x, y)).$$

From equation (8.3) we obtain that the power needed to satisfy a throughput $\bar{\theta}(x, y)$ will be $\theta^{\mathrm{RR}}(x, y) \geq \bar{\theta}(x, y)$, *i.e.*,

$$P_i(x, y) \geq \frac{\sigma^2}{h_i(x, y)} (2^{N_i \bar{\theta}(x,y)} - 1). \tag{8.4}$$

As our objective function is to minimize the total power of the network, the constraint will be reached, and from equation (8.2) we obtain

$$P_i(x, y) = \sigma^2 (2^{N_i \bar{\theta}(x,y)} - 1)(R^2 + d_i^2(x, y))^{\mathcal{P}/2}. \tag{8.5}$$

From last equation (8.5) we can observe that:

- If the quantity of mobile terminals increases inside the cell, it will need to transmit more power to each of the mobile terminals. The reason is that the base station is dividing each time-slot into mini-slots with respect to the number of the mobiles within its cell.

- The function $(R^2 + d_i^2(x,y))^{\mathcal{P}/2}$ on the right hand side give us the dependence of the power with respect to the distance between the base station and the mobile terminal located at position (x,y).

The problem that we are trying to solve deals with the optimal mobile association in order to minimize the total power of the network. Then the problem, that we denote (RR), reads

$$\text{(RR)} \quad \underset{C_i}{\text{Min}} \sum_{i=1}^{K} \iint_{C_i} P_i(x,y)\lambda(x,y)\,dx\,dy.$$

where $\lambda(x,y)$ is the function of distribution of the users. From equation (8.5) we obtain that in order to minimize the total power of the network using the round robin scheduling policy the problem, that we denote (RR), reads

$$\underset{C_i}{\text{Min}} \sum_{i=1}^{K} \iint_{C_i} \sigma^2(R^2 + d_i(x,y)^2)^{\mathcal{P}/2}(2^{N_i\Theta(x,y)} - 1)\lambda(x,y)\,dx\,dy.$$

We will solve this problem in section 8.4.

Formulation for the fairness problem

The general formulation for the problem of maximization of a function of the throughput given the constraint on the maximal power used admits a generalized α-fairness formulation given by:

$$\text{Max} \sum_{i=1}^{K} \iint_{C_i} \frac{1}{1-\alpha}[f(\theta_i(x,y))^{1-\alpha} - 1]\lambda(x,y)\,dx\,dy$$

where we can identify different problems for different values of α:

- $\alpha = 0$ maximization of throughput problem
- $\alpha \to 1$ proportional fairness (a uniform case of Nash bargaining)
- $\alpha = 2$ delay minimization
- $\alpha \to +\infty$ max-min fairness (maximize the minimum throughput that a user can have).

Since in our setting the problem is different since we are minimizing the total power on the network given the constraint of a minimum level of throughput we define the following formulation, that we call generalized γ-fairness:

$$\text{Min} \sum_{i=1}^{K} \iint_{C_i} \frac{1}{\gamma-1}[f(P_i(x,y)^{\gamma-1} - 1]\lambda(x,y)\,dx\,dy$$

where we can also identify different problems for different values of γ:

- $\gamma = 0$ maximization of the inverse of power (energy efficiency maximization)
- $\gamma \to 1$ proportional fairness
- $\gamma = 2$ minimization of total power
- $\gamma \to +\infty$ min-max fairness[1] (to minimize the maximum power per BS).

This problem is studied in section 8.5.

Rate fair allocation policy

- User optimization

In the round robin scheduling policy each base station BS_i modifies the power sent to mobile terminals with different channel gains in order to satisfy a throughput of $\Theta(x, y)$ for each mobile located at position (x, y). Instead, in the rate fair allocation policy each base station BS_i will maintain a constant power P_i sent to mobile terminals within its cell, *i.e.*,

$$P_i(x, y) = P_i \quad \text{for each } (x, y) \in C_i, \tag{8.6}$$

but it will modify the fraction of time allotted to the mobile terminals set in a way such that the average transmission rate to each mobile terminal with different channel gain is the same $\Theta(x, y)$ for each mobile located at position (x, y).

Let r_i be the fixed rate of mobile terminals located inside cell C_i. Following the rate fair allocation policy, the fraction of time that a mobile terminal at position $(x, y) \in C_i$ receives positive throughput will be

$$\frac{r_i}{\mathrm{SNR}_i(x, y)}.$$

Then the fixed rate r_i is the solution to the equation

$$\iint_{C_i} \frac{r_i}{\mathrm{SNR}_i(x_0, y_0)} \lambda(x, y)\, dx\, dy = \Theta := 2^{\bar{\theta}} - 1,$$

where $\bar{\theta}$ is the throughput to be satisfied. Then the rate

$$r_i = \Big(\iint_{C_i} \frac{1}{\mathrm{SNR}_i(x, y)} \lambda(x, y)\, dx\, dy \Big)^{-1} \Theta.$$

From equations (8.3) and (8.6) replacing the SNR we obtain

$$r_i = \Big(\iint_{C_i} \frac{\sigma^2}{P_i h_i(x, y)} \lambda(x, y)\, dx\, dy \Big)^{-1} \Theta,$$

[1]The min-max fairness is not well studied in the literature but one can map the max-min fairness studies into the min-max fairness for minimization problem. The convexity properties required becomes concavity, Schur convexity, sub-stochastic ordering, etc.

and from equation (8.2) we obtain

$$r_i = \Theta P_i \Big(\iint_{C_i} \sigma^2 (R^2 + d_i(x,y))^{\mathcal{P}/2} \lambda(x,y)\, dx\, dy \Big)^{-1}, \tag{8.7}$$

We seek for an equilibrium in the game in which each mobile terminal chooses to which base station is going to be served. Similar notion of equilibrium has been studied in the context of large number of small players in transportation by Wardrop [2].

Definition.- The Wardrop equilibrium is given by:

$$\text{If } \iint_{C_i} \lambda(x,y)\, dx\, dy > 0, \text{ then } r_i = \max_{1 \le j \le K} r_j(C_j), \tag{8.8a}$$

$$\text{and if } \iint_{C_i} \lambda(x,y)\, dx\, dy = 0, \text{ then } r_i \le \max_{1 \le j \le K} r_j(C_j).$$

As in our case we consider that the area of each cell is non-zero and the distribution of the mobile terminals within each cell is positive, then the equilibrium situation will be given by

$$r_1 = r_2 = \ldots = r_K.$$

To understand this equilibrium situation, consider as an example the case of two base stations BS_i and BS_j. Assume that one of the base stations BS_i offer more rate than the other base station BS_j, then the mobiles served by BS_j will have an incentive to be served by base station BS_i. Notice that the terms inside the integral of equation (8.7) are all positive. Then the rate transmitted from base station depends inversely on the quantity of mobiles inside the cell. It depends on the quantity of mobiles through the size of the cell C_i and through the density of mobiles inside the cell $\lambda(x,y)$. As more mobile terminals will try to connect to the base station BS_i the rate will diminish until arrive to the equilibrium where both base stations will offer the same rate.

Let us denote by r to the rate offered by the base station at equilibrium, *i.e.*,

$$r := r_1 = r_2 = \ldots = r_K.$$

Then from equation (8.7)

$$P_i(C_i) = \frac{r}{\Theta} \iint_{C_i} \sigma^2 (R^2 + d_i^2(x,y))^{\mathcal{P}/2} \lambda(x,y)\, dx\, dy. \tag{8.9}$$

We want to choose the optimal mobile assignment in order to minimize the total power of the network under the constraint that the mobile terminals have an average throughput of θ, *i.e.*,

$$\operatorname*{Min}_{C_i} \sum_{i=1}^{K} P_i(C_i) \tag{8.10}$$

Then our problem reads

$$\text{(RF)} \quad \underset{C_i}{\text{Min}} \sum_{i=1}^{K} \iint_{C_i} \sigma^2 (R^2 + d_i^2(x,y))^{\mathcal{P}/2} \lambda(x,y)\, dx\, dy.$$

We will solve this problem in section 8.6.

8.2.2 Uplink Case

Consider the SINR density given by base station BS_i located at y as in Altman *et al.* [99]

$$\text{SINR}_i(x) = \frac{[R^2 + (y-x)^2]^{-\mathcal{P}/2}}{\int_D (R^2 + (y-z)^2)^{-\mathcal{P}/2}\, dz + \sigma^2}$$

In this case, the authors of [99] considered a uniform distribution of mobile terminals and a constant power. We generalize their setting by considering a density of mobile terminals $\lambda(x)$ and a power given by $P_i(x)$ in the one dimensional case. Then the problem reads

$$\text{SINR}_i(x) = \frac{P_i(x)[R^2 + (y-x)^2]^{-\mathcal{P}/2}}{\int_D P_i(z)(R^2 + (y-z)^2)^{-\mathcal{P}/2} \lambda(z)\, dz + \sigma^2}\, dx$$

This can be generalized to the two dimensional case

$$\text{SINR}_i(x,y) = \frac{P_i(x,y)(R^2 + d_i(x,y)^2)^{-\mathcal{P}/2}}{P_{\text{total}} + \sigma^2}\, dx,$$

where

$$P_{\text{total}} := \iint_D (R^2 + d_i(x,y)^2)^{-\mathcal{P}/2} \lambda(x,y)\, dx\, dy.$$

As we want to guarantee an average SNR of $\Theta(x,y)$ to a mobile located at position (x,y) this condition is written as

$$\frac{P_i(x,y)(R^2 + d_i(x,y)^2)^{-\mathcal{P}/2}}{P_{\text{total}} + \sigma^2}\, dx \geq \Theta(x,y).$$

Then as the constraint will be reached it follows that

$$P_i(x,y) = \Theta(x,y)(P_{\text{total}} + \sigma^2)(R^2 + d_i(x,y)^2)^{+\mathcal{P}/2}.$$

And then our problem reads

$$\underset{C_i}{\text{Min}} \sum_{i=1}^{K} \iint_{C_i} P_i(x,y) \lambda(x,y) \Theta(x,y)\, dx\, dy$$

We denote this problem as (UL) and replacing the power is written as

$$\underset{C_i}{\mathrm{Min}} \sum_{i=1}^{K} \iint_{C_i} (P_{\mathrm{total}} + \sigma^2)(R^2 + d_i(x,y)^2)^{\mathcal{P}/2} \lambda(x,y) \Theta(x,y) \, dx \, dy$$

which is similar except by a constant to our problem (RF).

In order to solve the problem of the round robin scheduling policy (RR), the rate fair allocation policy (RF), and the uplink case (UL) we will make use of Optimal Transport Theory. a theory that has prove to be useful on many economical context [90], [91], [92], as well as in the road traffic community [93], but to the best of the authors knowledge it has never been used in the telecommunication community.

8.3 Basics in optimal transport theory

The theory of mass transportation, also called optimal transport theory, goes back to the original works by Monge in 1781 [100], and later in 1942 by Kantorovich [4].

The work by Brenier [101] has renewed the interest for the subject and since then many articles have been published in this topic (see [102] and references therein).

The original Monge's problem can be interpreted as the question: "How do you best move given piles of sand to fill up given holes of the same total volume?". The general mathematical framework to deal with this problem is a little technical but we encourage to jump the details and to focus on the main ideas.

We first consider a grid area network D in the one-dimensional case. As an example, the function $f(t)$ will represent the proportion of how much sand is located at t and we denote

$$d\mu(t) := f(t) \, dt.$$

The function $g(s)$ will represent the proportion of how much sand can be piled at location s and we denote

$$d\nu(s) := g(s) \, ds.$$

The function T (called transport map) is the function that transfers sand from location s to location t. The condition of conservation that the sand transferred is equal to the sand received gives

$$\int_A g(y) \, dy = \int_{\{x \,:\, T(x) \in A\}} f(x) \, dx$$

and we denote this condition $T\#\mu = \nu$.

The original problem was to move piles of sand to holes, Monge's problem considered that the cost of moving sand from position x to position y depends on the distance $c(|x-y|)$. Then the cost of moving sand from position x through T to its image position $T(x)$ will be $c(|x - T(x)|)$. We consider the total cost over D. Then Monge's problem is

$$\mathrm{Min} \int_{\mathrm{D}} c(|x - T(x)|) \, f(x) \, dx \quad \text{such that} \quad T\#\mu = \nu.$$

The main difficulty in solving Monge's problem is the highly non-linear structure of the objective function. As an example, consider the domain $\mathrm{D} = [0, 2]$, the throughput from the base stations located at position 1 to the mobile terminals denoted $\mu = \delta_1$ and the throughput of two mobile terminals demanded to the base stations located at positions 0 and 2, denoted $\nu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_2$. According to the formulation given by Monge, there is no splitting of throughput so this problem doesn't have a transport map (see Fig. 8.1). We pointed out the limitations of Monge's problem that motivated Kantorovich to consider another modeling of this problem in [4].

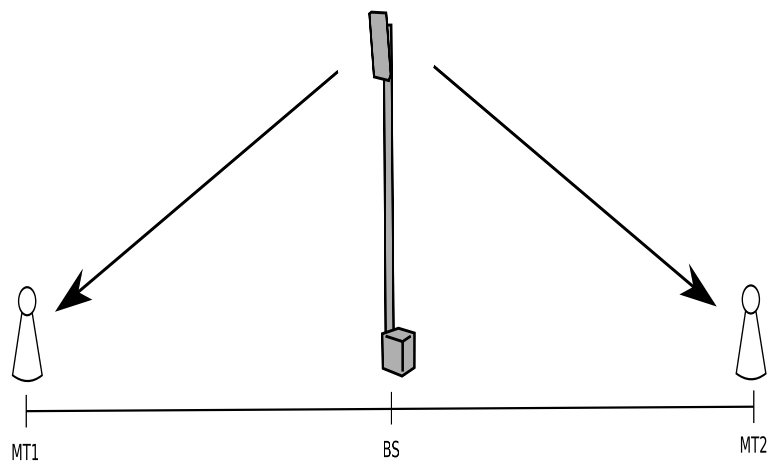


Figure 8.1: Monge's problem can not model a simple scenario of two mobile terminals and one base station. Kantorovich's problem however can model very general scenarios.

Kantorovich noticed that the problem of transportation from one location to another can be seen as "graphs" (called transport plans) of functions in the product space (See Fig. 8.2).

Then Kantorovich's problem is

$$\underset{\psi \in \Pi(\mu,\nu)}{\mathrm{Min}} \iint_{\mathrm{D}\times\mathrm{D}} c(x, y)\, d\psi(x, y)$$

where $\Pi(\mu, \nu) = \{\psi : \pi_1\#\psi = \mu \text{ and } \pi_2\#\psi = \nu\}$ is denoted the ensemble of transport plans ψ, $\pi_1(x, y)$ stands for the projection on the first axis x, and $\pi_2(x, y)$ stands for the projection on the second axis y.

The relationship between Monge and Kantorovich problems is that every transport map T of Monge's problem determines a transport plan $\psi = (\mathrm{Id} \times T)\#\mu$ in Kantorovich's problem with the same cost (Id denotes the identity). However, Kantorovich's problem consider more functions than the ones coming from Monge's problem, so we can choose from a bigger set $\Pi(\mu, \nu)$.

We denote when it exists

$$M_p(\mu, \nu) := \left(\underset{T\#\mu=\nu}{\mathrm{Min}} \int_{\mathrm{D}} |x - T(x)|^p\, f(x)\, dx \right)^{1/p}$$

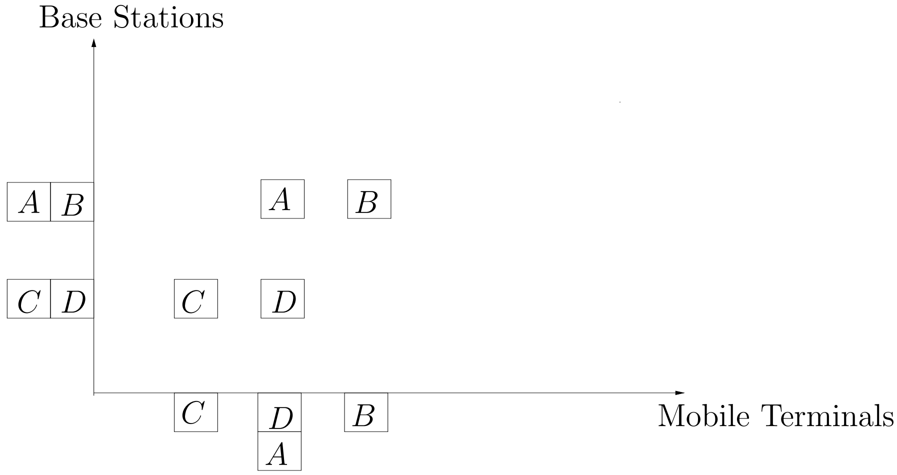


Figure 8.2: Kantorovich considered "graphs" where the projection in the first axis coincide with the mobile terminal position ($\mathrm{MT}_1 = 3.5$, $\mathrm{MT}_2 = 5$ and $\mathrm{MT}_3 = 6.5$) and the second axis coincides with the base station position ($\mathrm{BS}_1 = 4$ and $\mathrm{BS}_2 = 6$).

$$\text{and} \quad W_p(\mu,\nu) := \left(\underset{\psi\in\Pi(\mu,\nu)}{\mathrm{Min}} \iint_{\mathrm{D}\times\mathrm{D}} |x-y|^p \, d\psi(x,y) \right)^{1/p}.$$

We are now ready to give a result on existence and uniqueness of the transport plan.

Theorem 8.3.1 (Existence and uniqueness) *Consider the cost function $c(|x-y|) = |x-y|^p$. Let μ and ν be probability measures in* D *and fix $p \geq 1$. We assume that μ can be written*[2] *as $d\mu = f(x)\,dx$. Then the optimal value of Monge's problem coincides with the optimal value of Kantorovich's problem, i.e., $M_p(\mu,\nu) = W_p(\mu,\nu)$ and there exists an optimal transport map from μ to ν, which is also unique almost everywhere if $p > 1$.*

This result is very difficult to obtain and it has been proved only recently (see [101] for the case $p = 2$, and the references at [102] for the other cases).

The case that we are interested in can be characterized because the image of the transport plan is a discrete finite set.

Since the problem is a linear optimization problem under linear constraints we look at the dual formulation of Kantorovich's relaxation problem:

Theorem 8.3.2 (Dual formulation) *For μ and ν probability measures in* D*, the following equality holds:*

$$W_p^p(\mu,\nu) = \sup \left(\int_{\mathrm{D}} u \; d\mu + \int_{\mathrm{D}} v \; d\nu \right) \quad \textit{such that}$$

$$\begin{cases} u \in L^1_\mu, v \in L^1_\nu \\ u(x) + v(y) \leq |x-y|^p \quad \mu \textit{ and } \nu \textit{ almost everywhere} \end{cases}$$

[2]The exact condition is that μ is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^d$ where d is the dimension of the space.

$$\textit{where} \quad u \in L^1_\mu \quad \textit{means that} \quad \int_D u(x) f(x)\, dx < +\infty$$

and similarly for $v \in L^1_\nu$. *Moreover, there exists an optimal pair* (u, v) *for this dual formulation.*

Remark 8.3.1 *In the particular case when* $\nu = \sum_{i \in \mathbb{N}} b_i \delta_{y_i}$ *is a sum of Dirac measures, the dual formulation reads*

$$W_p^p \left(\mu, \sum_{i \in \mathbb{N}} b_i \delta_{y_i} \right) = \sup \left\{ \int_{\mathrm{D}} u \, d\mu + \sum_{i \in \mathbb{N}} b_i v(y_i) \right\}$$

$$\begin{cases} u \in L^1_\mu(\mathrm{D}), v \in L^1_\nu(\mathrm{D}) \\ u(x) + v(y_i) \leq |x - y_i|^p \quad \textit{for } \mu\textit{-a.e. } x \textit{ and every } i \in \mathbb{N}. \end{cases}$$

Remark 8.3.2 *In the particular case when* μ *can be written as* $d\mu = f(x)\, dx$ *and* $\nu = \sum_{i \in \mathbb{N}} b_i \delta_{y_i}$ *any transport map* T *is associated to a partition* $(B_i)_{i \in \mathbb{N}}$ *of* D *satisfying* $\mu(B_i) = b_i$. *As* $(B_i)_{i \in \mathbb{N}}$ *is a partition,* x *belongs to some element of the partition* B_j *and then we associate it to* y_j, *i.e.,* $T(x) = y_j$.

Thanks to optimal transport theory we are able to characterize the partitions on very general settings. For doing so, consider locations $(x_1, y_1) \ldots, (x_K, y_K)$, the Euclidean distance $d_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$, and F a continuous function.

Theorem 8.3.3 *Consider the problem* (P1)

$$\operatorname*{Min}_{C_i} \sum_{i=1}^{K} \iint_{C_i} \left[F(d_i(x, y)) + s_i \left(\iint_{C_i} \lambda(\omega, z)\, d\omega\, dz \right) \right] \lambda(x, y)\, dx\ dy,$$

where C_i *is the cell partition of* D. *Suppose that* s_i *are continuously differentiable, non-decreasing, and convex functions. The problem* (P1) *admits a solution that verifies*

$$(\mathrm{S1}) \begin{cases} C_i = & \{x : F(d_i(x, y)) + s_i(N_i) + N_i \cdot s_i'(N_i) \leq \\ & \leq F(d_j(x, y)) + s_j(N_j) + N_j \cdot s_j'(N_j)\} \\ N_i = & \iint_{C_i} \lambda(\omega, z)\, d\omega\, dz. \end{cases}$$

Theorem 8.3.4 *Consider the problem* (P2)

$$\operatorname*{Min}_{C_i} \sum_{i=1}^{K} \iint_{C_i} \left[F(d_i(x, y)) \cdot m_i \left(\iint_{C_i} \lambda(\omega, z)\, d\omega\, dz \right) \right] \lambda(x, y)\, dx\, dy$$

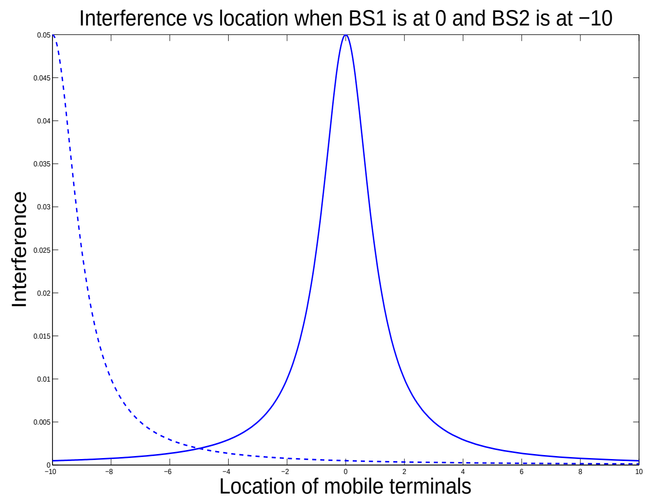


Figure 8.3: Interference as a function of location of mobile terminals when BS_1 is at position 0 (solid line) and BS_2 at -10 (dashed line).

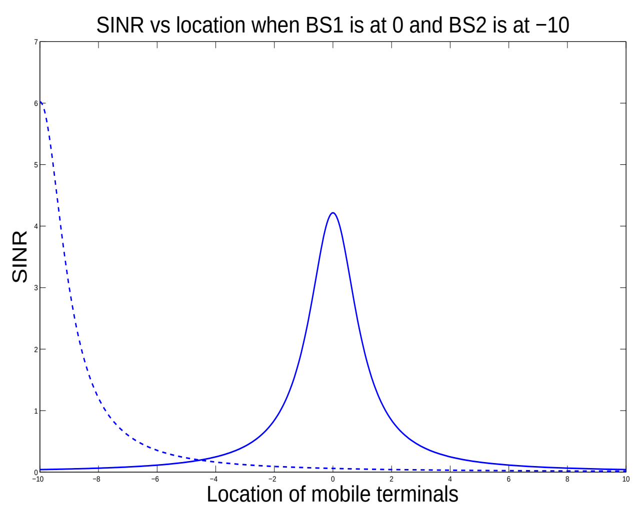


Figure 8.4: SINR as a function of location of mobile terminals when BS_1 is at position 0 (solid line) and BS_2 at -10 (dashed line).

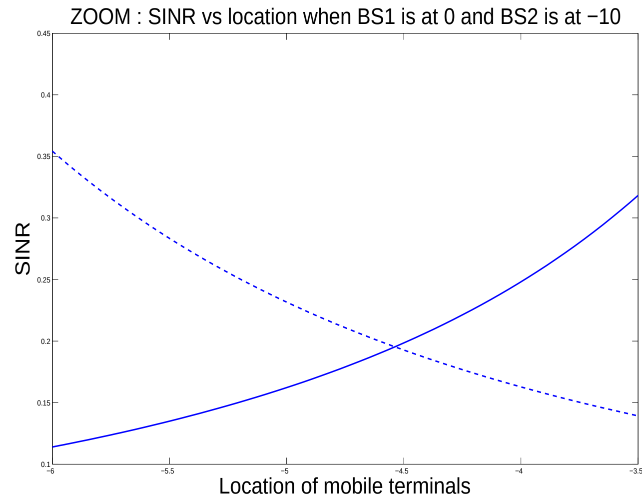


Figure 8.5: Zoom of the SINR as a function of the location of mobile terminals when BS_1 is at position 0 (solid line) and BS_2 is at position -10 (dashed line). The best equilibrium is $eq_1 = -4.68$ with SINR value of 0.0025.

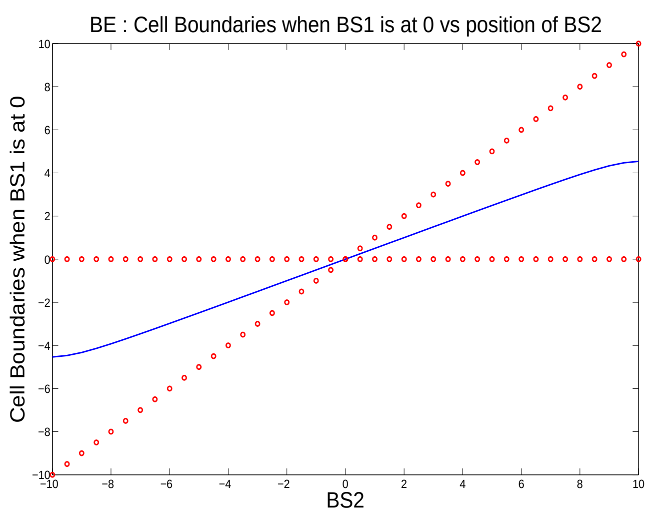


Figure 8.6: Best Equilibrium: Thresholds determining the cell boundaries (vertical axis) as a function of the location of BS_2 for BS_1 at position 0.

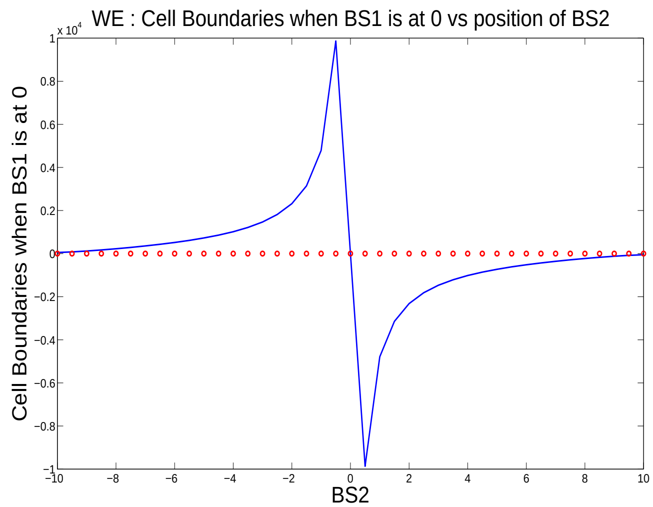


Figure 8.7: Worst Equilibrium: Thresholds determining the cell boundaries (vertical axis) that give the worst equilibrium in terms of the SINR as a function of the location of BS_2 for BS_1 at position 0.

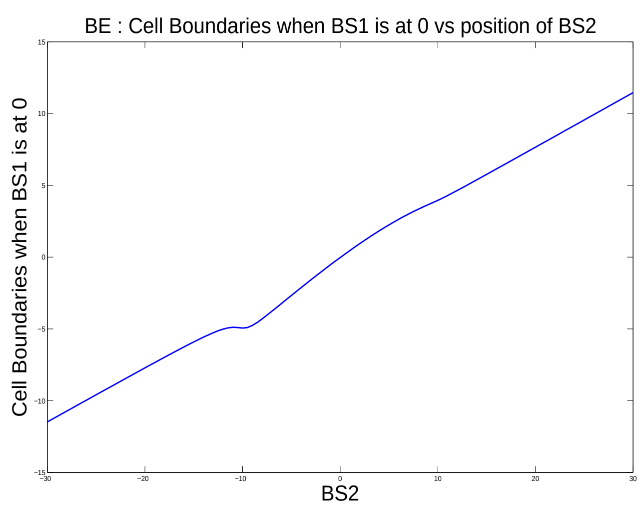


Figure 8.8: Non-Homogeneous case: Thresholds determining the cell boundaries (vertical axis) of the best equilibrium in terms of the SINR as a function of the location of BS_2 for BS_1 at position 0 when we consider a distribution given by $\lambda(x) = (L - x)/2L^2$.

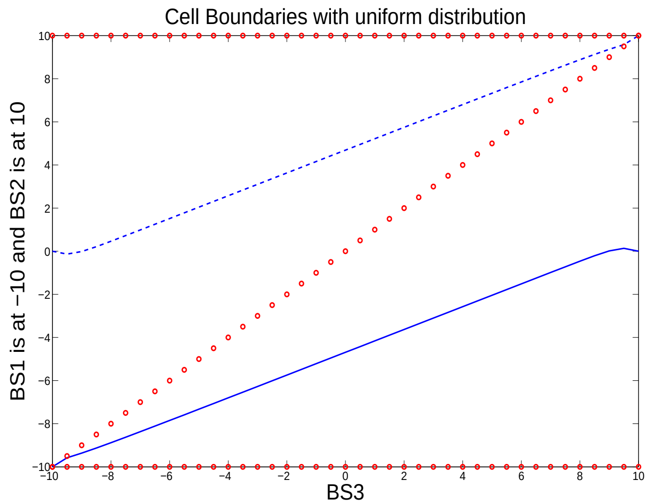


Figure 8.9: Several BSs: Threshold determining the cell boundaries as a function of the location of BS_3 for $BS_1 = -10$ and $BS_2 = 10$

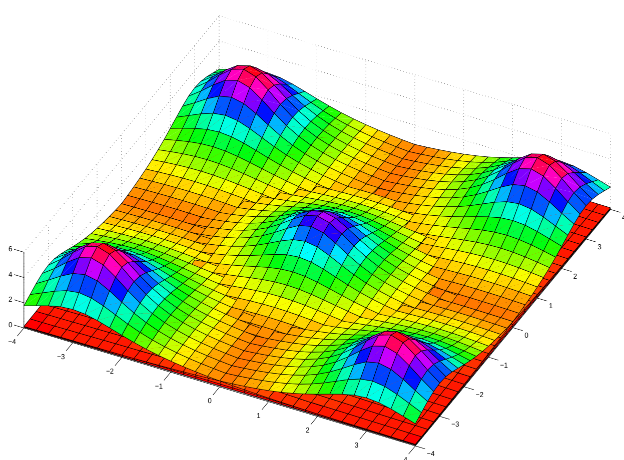


Figure 8.10: 2D case: Cell boundaries of the best equilibrium with uniform distribution of users.

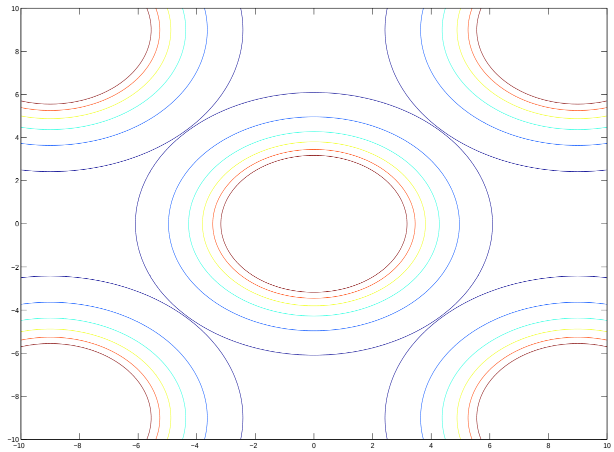


Figure 8.11: 2D case: Cell contours of the best equilibrium with uniform distribution of users.

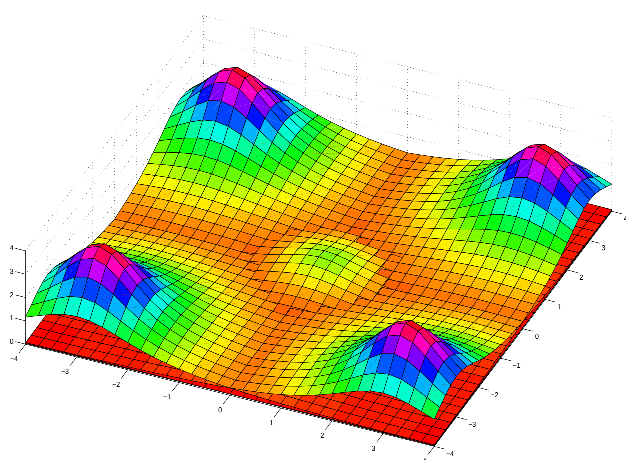


Figure 8.12: 2D Non-Homogeneous case: Cell boundaries of the best equilibrium with non-homogeneous distribution of users.

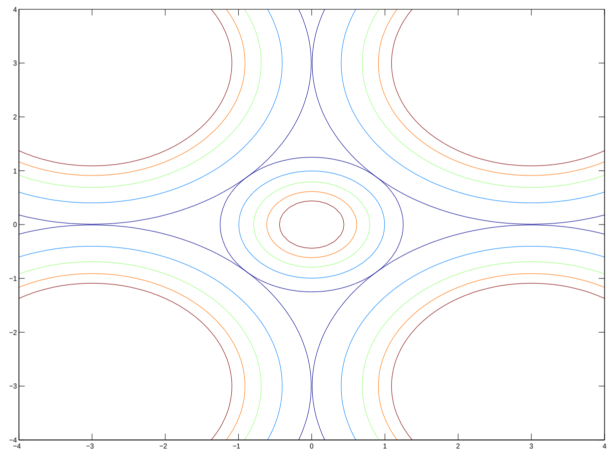


Figure 8.13: 2D Non-Homogeneous : Cell contours of the best equilibrium with non-homogeneous distribution of users.

where C_i *is the cell partition of* D. *Suppose that* m_i *are derivable. The problem* (P2) *admits a solution that verifies*

$$(\mathrm{S2})\begin{cases} C_i = & \{x \,:\, m_i(N_i)F(d_i(x,y))\,\lambda(x,y) + U_i(x,y) \leq \\ & \leq m_j(N_j)F(d_j(x,y))\,\lambda(x,y) + U_j(x,y)\} \\ U_i = & m_i'(N_i)\iint_{C_i} F(d_i(x,y))\lambda(x,y)\,dx\,dy \\ N_i = & \iint_{C_i}\lambda(\omega,z)\,d\omega\,dz. \end{cases}$$

Notice that in problem (P1) if the functions $s_i \equiv 0$ the solution of the system (S1) becomes the well known Voronoi cells. In problem (P2) if we have that the functions $h_i \equiv 1$ we find again the Voronoi cells. However in all the other cases the Voronoi configuration is not optimal.

8.4 Round robin scheduling policy

We assume that a service provider wants to minimize the total power of the network while maintaining a certain average throughput of θ to each mobile terminal of the system using the round robin scheduling policy given by problem (RR)

$$\operatorname*{Min}_{C_i}\sum_{i=1}^{K}\iint_{C_i}\sigma^2(R^2+d_i(x,y)^2)^{\mathcal{P}/2}(2^{N_i\theta}-1)\lambda(x,y)\,dx\,dy.$$

We see that this problem is an optimal transportation problem (P1) with cost function given by

$$F(d_i(x,y)) = \sigma^2(R^2+d_i(x,y)^2)^{\mathcal{P}/2}$$

$$m_i(x,y) = (2^{N_i\theta} - 1)$$

Proposition.- There exist a unique optimum given by

$$\begin{aligned} C_i = \Big\{ x \in \mathcal{D} \ &: d_i(x_0,y_0)^p + h_i(N_i) + N_i h_i'(N_i) \\ &\le d[(x_0,y_0),(x_j,y_j)]^p + k_j(N_j) + N_j k_j'(N_j) \quad \forall j \neq i \Big\} \\ N_i = \iint_{C_i} &\lambda(x_0,y_0)\,dx_0\,dy_0 \end{aligned}$$

Proof.- See Appendix A.

Example.- Consider a network of $N = 2500$ mobile terminals distributed according to $\lambda(x)$ in $[0, L]$ (for example, with $L = 5.6$ miles for WiMAX radius cell). We consider two base stations at position $\mathrm{BS}_1 = 0$ and $\mathrm{BS}_2 = L$ and the high of the base stations is scaled to be $R = 1$. Then the system of equations is reduced to find x such that:

$$\begin{aligned} (2^{N_1\theta} - 1)(1 + x^2)\lambda(x) + 2^{N_1\theta}\theta \log 2 \left[x + \frac{x^3}{3} \right] \\ = (2^{N_2\theta} - 1)(1 + (1-x)^2)\lambda(x) + \\ 2^{N_2\theta}\theta \log 2 \left[\frac{4}{3} - 2x + x^2 - \frac{x^3}{3} \right] \end{aligned}$$

This is a fixed point equation on x since N_1, N_2 and λ depend on x. When mobile terminals are distributed uniformly, the optimal solution is given by $[0, 1/2)$ and $[1/2, 1]$, which is the case of Voronoi cells and the number of mobile terminals connected to each base station is equal and given by $N_1 = N_2 = 1250$. However when the distribution of mobile terminals is increasingly more concentrated at location L, given by $\lambda(x) = 2x$, the optimal solution is given by $[0, q)$ and $[q, 1]$ with $q = 0.6027$ and the quantity of mobile terminals connecting to BS_1 is equal to $N_1 = 908$ and the quantity of mobile terminals connecting to BS_2 is equal to $N_2 = 1592$ (See Fig. 8.14).

8.5 Fairness problem

As we mention in section 8.2 the solution given by previous section 8.4 is optimal but may not be fair to all the mobile terminals since it will give higher throughput to the mobile terminals that are near the base stations.

To deal with this problem we considered the fairness problem given by

$$\begin{aligned} \mathrm{Min} \sum_{i=1}^{K} \iint_{C_i} \frac{1}{\gamma - 1} \left(\sigma^2 (R^2 + d_i(x,y)^2)^{\mathcal{P}/2} \right)^{\gamma-1} \\ (2^{N_i\theta} - 1)^{\gamma-1} \lambda(x,y)\,dx\,dy. \end{aligned}$$

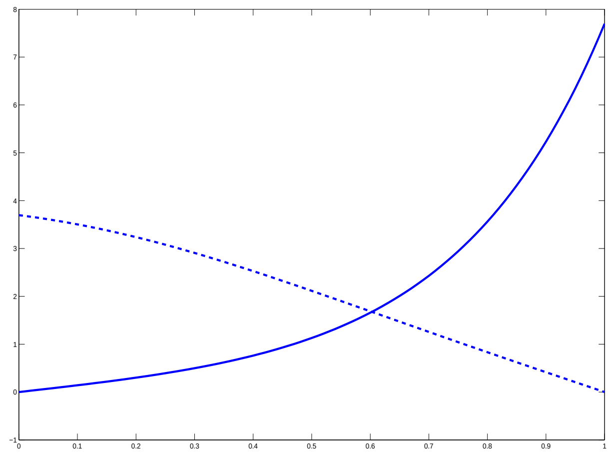


Figure 8.14: Example: equilibrium when then distribution of mobile terminals is given by $\lambda(x) = 2x$ in the interval $[0, L]$. and the positions of the base stations are $\mathrm{BS}_1 = 0$ and $\mathrm{BS}_2 = L$.

As we can see this is also an optimal transportation problem (P1) where the functions considered in this setting are given by

$$F(d_i(x,y)) = \frac{1}{\gamma - 1}\left(\sigma^2(R^2 + d_i(x,y)^2)^{\mathcal{P}/2}\right)^{\gamma-1}$$

$$m_i(x,y) = (2^{N_i\theta} - 1)^{\gamma-1}$$

Using Theorem 8.3.3 we are able to characterize the optimal cells for any γ considered.

8.6 Rate fair allocation policy

In this framework we give the possibility to mobile terminals to connect to the base station they prefer in order to minimize their power cost function while maintaining an average throughput of θ. This is the reason why we denote this type of network as *hybrid network*.

As we saw this problem is equivalent to

$$\text{(RF)} \quad \underset{C_i}{\mathrm{Min}} \sum_{i=1}^{K} \iint_{C_i} \sigma^2(R^2 + d_i^2(x,y))^{\mathcal{P}/2} \lambda(x,y)\, dx\, dy.$$

Notice that this problem is equivalent to (P1) where the functions $s_i \equiv 1$ The problem has then a solution given by

Proposition.- There exist a unique optimum given by

$$C_i = \Big\{x \in \mathcal{D} \ : \sigma^2(R^2 + d_i^2(x_0, y_0))^{\mathcal{P}/2}$$
$$\leq \sigma^2(R^2 + d_j^2(x_0, y_0))^{\mathcal{P}/2} \quad \forall j \neq i\Big\}$$
$$N_i = \iint_{C_i} \lambda(x_0, y_0)\, dx_0\, dy_0$$

which is the Voronoi cells.

8.7 Uplink case

Penalization function As an illustration example, suppose that on the network D $= [0, 1]$ there are two base stations at coordinates $x_1 = 1/4$ and $x_2 = 3/4$. Assume that mobile terminals are uniformly distributed, and consider the case when $\mathcal{P} = 2$.

Suppose the first base station can handle more downlink demand than the second one, as for example the first base station uses a IEEE 802.16 (WiMAX) technology while the second one uses UMTS technology, so that the penalization cost are

$$h_1(t) = t \quad \text{and} \quad h_2(t) = (1 + \varepsilon)t.$$

Then the optimum cell configuration (C_1^*, C_2^*) is given by

$$C_1^* = [0, \lambda_\varepsilon^*[, \quad C_2^* =]\lambda_\varepsilon^*, 1] \quad \text{with} \quad \lambda_\varepsilon^* = \frac{1}{2} + \frac{\varepsilon}{5 + 2\varepsilon},$$

whereas the equilibrium cell configuration (C_1^E, C_2^E) will be $C_1^E = [0, \lambda_\varepsilon^E[, \quad C_2^E =]\lambda_\varepsilon^E, 1]$ with

$$\lambda_\varepsilon^E = \frac{1}{2} + \frac{\varepsilon}{6 + 2\varepsilon} \leq \lambda_\varepsilon^*.$$

8.8 Validation of our theoretical model

8.8.1 One-dimensional case: uniform distribution of users

We first consider the one-dimensional case and we consider a uniform distribution of users in the interval $[-L, L]$. We set $L = 10$ and the noise parameter $\sigma = 0.3$. We fix one base station BS_2 at position 0 and we move the other base station BS_1. We consider the path loss exponent of $\mathcal{P} = 2$.

In the SINR-association game we found two pure equilibria: the best equilibrium at position $\text{eq}_1 = -4.68$ with SINR value of 2.5×10^{-3} and the worst equilibrium at position $\text{eq}_2 = 78.69$ with SINR value of 1.4769×10^{-9}. It is known than any other mixed equilibrium will give lower values of SINR. From now on we will only be interested in the best equilibrium. See Fig. 8.6 and Fig. 8.7.

We found that even in the one-dimensional case, the results of [99] are not-valid, the cells are convex and monotone inside the network.

8.8.2 One-dimensional case: non-uniform distribution of users

In this case we consider a non-uniform distribution of users $\lambda(x) = (L - x)/L^2$ under the same setting as in 8.8.1. we found again that the cells are convex and monotone inside the network.

8.8.3 Two-dimensional case: uniform distribution of users

We consider the two-dimensional case and we consider a uniform distribution of users in the square $[-L, L] \times [-L, L]$. We set $L = 10$ and the noise parameter $\sigma = 0.3$. We set five base stations at positions $\mathrm{BS}_1 = (-L+1, -L+1)$, $\mathrm{BS}_2 = (L-1, -L+1)$, $\mathrm{BS}_3 = (-L+1, L-1)$, $\mathrm{BS}_4 = (L-1, L-1)$, and $\mathrm{BS}_5 = 0$. Numerically we observe again that the cells are convex and monotone inside the domain. See Fig. 8.11

8.8.4 Two-dimensional case: non-uniform distribution of users

We consider the two-dimensional case and this time we consider a non-uniform distribution of users in the square $[-L, L] \times [-L, L]$ given by $\lambda(x, y) = (L^2 - (x^2 + y^2))/K$ where K is a normalization factor. This situation can be interpreted as the situation when mobile terminals are more concentrated in the center and less concentrated in suburban areas as in Paris, New York or London. We observe that the cell size of the base station BS_5 at the center is smaller than the others at the suburban areas. This can be explained by the fact that as the density of users is more concentrated in the center the interference is greater in the center than in the suburban areas and then the SINR is smaller in the center. However the quantity of users is greater than in the suburban areas.

See Fig. 8.13.

8.9 Conclusions

We have proposed a new approach using optimal transport theory for mobile association and we have been able to completely characterize this mobile association under different policies in both uplink and downlink cases.

Conclusions and Perspectives

Conclusions

The growing number of wireless systems and wireless devices present many challenges for industry and academia in the planning and analysis of wireless networks. In this manuscript, we focus on the modeling and analysis of massively dense wireless networks, in particular, massively dense ad hoc networks and massively dense cellular systems. We study the continuum modeling approach, which is useful for the initial phase of deployment of the network, as well as to analyze broad-scale regional studies of the network. In this type of studies, the focus is on the general trend and pattern of the transmission distribution through the network.

The modeling of congestion-aware routing problems for a network can be classified in the discrete modeling approach and the continuum modeling approach:

- In the discrete modeling approach, the network is modeled as a graph, where each wireless link connecting two nodes in the network is modeled separately an the demand is assumed to be concentrated at some part of the nodes. This modeling approach is commonly adopted for the detailed planning and analysis of the network.

- In the continuum modeling approach, we are interested on the macroscopic behavior of the network. This macroscopic analysis is not as detailed as the discrete modeling approach, but nevertheless, it contains enough information to allow meaningful results.

In the first part of the thesis, we have investigated the routing optimization problem in massively dense ad hoc networks, where we have considered a generic cost function, that can take into account different metrics such as the congestion of the network, the quantity of relay nodes needed to maintain a certain throughput, or metrics related to the energy consumption of the network. Following a similar approach to the work of Nash [1] and Wardrop [2], we have defined for massively dense networks two principles of network optimization, which we denoted, respectively, user-optimization and system-optimization:

- The first principle takes into account the situation in which each user select its routes from origins to destinations, in order to minimize its own cost. Then, in an equilibrium situation, the cost of all routes actually used between an origin/destination pair are equal. This cost turns out to be less than the cost that would be experienced by a single user on any of the unused routes. In road-traffic theory, the user-optimized solution is also referred to as the traffic network equilibrium.

- The second principle reflects the situation in which there is a central controller which decide the routes and the traffic flows in an optimal manner from origins to destinations to minimize the total cost of the network.

Beckmann, McGuire, and Winsten [7] were the first to provide a rigorous mathematical formulation of the conditions set forth by Wardrop's first principle in the context of certain link cost functions, which were increasing functions of the flows on the links. In particular, they demonstrated that the optimality conditions in the form of Karush-Kuhn-Tucker [8, 9] conditions of an appropriately constructed mathematical programming/optimization problem coinciding with Wardrop's first principle. We have shown that a similar analysis can be done for the continuum modeling approach. With different cost functions, we were able to formulate and solve the routing problem for the user- and system-optimization problem in two different contexts: for directional antennas and omnidirectional antennas. We have also found a simple characterization of the minimum cost paths by extensive use of Green's theorem in directional antennas.

In many situations, the optimal solution of the problem, in the user- as well as in the system-optimization problem, is characterized by a partial differential equation. We propose the numerical analysis of this equations by finite elements method which have allow us to give bounds in the variation of the solution with respect to the variation of the data. When we allow mobility of the origin and destination nodes, we are able to found the optimal quantity of relay nodes needed to transmit a certain quantity of data.

In the second part, our focus was on cellular networks, where we investigated the capacity of Network MIMO systems and MIMO broadcast channels as well as the mobile association problem in cellular networks.

In Network MIMO systems and MIMO broadcast channels, we have shown that, even when the channel offers an infinite number of degrees of freedom, the capacity is mainly limited by the ratio between the size of the antenna array at the base station and the mobile terminals and the wavelength of the signal.

For the mobile association problem, we were able to provide quality of service constraints while minimizing the total power of the network for the continuum modeling approach. We have solved in this context the user- and system-optimization problem under different policies and different distribution of the users in the network.

In conclusion, we have provided elements of analysis in both user-optimization and system-optimization networks. We have focus on the continuum modeling approach instead of the discrete modeling approach. Both approaches are not opposed, but complementary. However, most of the works in networks have been centered on the discrete modeling approach.

The continuum modeling approach has many advantages over the discrete approach in macroscopic studies on dense networks.

- First, it reduces the problem size for dense transportation networks. The problem size in the continuum model depends on the method that is adopted to approximate the modeling region, but not on the actual network itself. Because of that, an effective approximation method, such as the finite element method (FEM), can extensively reduce the size of the problem. This reduction in problem size saves computational time and memory.

- Second, less data is required to model the set-up in a continuum model. As continuum modeling can be characterized by a small number of spatial variables, it can be set-up with a much smaller amount of data than the discrete modeling approach, which requires data for all of the included links.

This makes the continuum model convenient for macroscopic studies in the initial phase of design since the collection of data in this phase is time consuming and labor intensive, and the resources to undertake it are generally not available, which means there is usually insufficient data on the system to set up a detailed model. Finally, the continuum modeling approach gives a better understanding of the global characteristics of a network.

Perspectives

There exists a number of open problems that can be seen inside this manuscript. However, from my point of view the most important problems that remain open are:

To break the boundary between the discrete and the continuous modeling approach In addressing this problem, we are thinking of constructing a consistent theoretical framework to address both problems and the convergence from the discrete problem into the continuous problem. In the same perspective, it would be of particular interest to study the user-optimization problem and its convergence from the discrete to the continuum approach.

Price of Anarchy on massively dense networks The price of anarchy is a common subject in routing. The price of anarchy is equal to the ratio of the utility obtained by selfish users to the utility they would obtain by the system-optimal solution. It measures the loss suffered by the system when there is no central controller. In general, to centralize the information and to take a global solution is expensive and in many ocassions impossible.

Bibliography

[1] J. F. Nash, "Noncooperatives games," *Annals of Mathematics*, vol. 54, pp. 286–298, 1951.

[2] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers*, vol. 1(II), pp. 325–378, 1952.

[3] F. L. Hitchcock, "The distribution of a product from several sources to numerous facilities," *Journal of Mathematical Physics*, vol. 20, pp. 224–230, 1941.

[4] L. V. Kantorovich, "On the translocation of masses," *Comptes Rendus (Doklady) de l'Academie des Sciences de l'URSS*, vol. 37, 1942.

[5] T. C. Koopmans, "Optimum utilization of the transportation system," in *Proceedings of the International Statistical Conference*, Washington DC, 1947.

[6] G. B. Dantzig, "Application to the simplex method to the transportation problem," T. C. Koopmans, Ed., New York, 1951, pp. 359–373.

[7] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. New Haven, Connecticut: Yale University Press, 1956.

[8] W. Karush, *Minima of functions of several variables with inequalities as side constraints*. Chicago, Illinois: Dept. of Mathematics, Univ. of Chicago, 1939.

[9] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, J. Neyman, Ed. Berkeley, California: University of California Press, 1951, pp. 481–492.

[10] S. C. Dafermos and F. T. Sparrow, "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards*, vol. 73B, pp. 91–118, 1969.

[11] M. J. Beckmann, "On the theory of traffic flows in networks," *Traffic Quarterly*, vol. 21, pp. 109–116, 1967.

[12] D. P. Bertsekas and R. Gallager, *Data Networks*. New Jersey: Prentice-Hall, Englewood Cliffs, 1987.

[13] D. Braess, "Ber ein paradoxen aus der verkehrsplanung," *Unternehmensforschung*, vol. 12, pp. 258–268, 1968.

[14] J. E. Cohen and F. P. Kelly, "A paradox of congestion in a queuing network," *Journal of Applied Probability*, vol. 27, pp. 30–34, 1990.

[15] S. Dafermos and A. Nagurney, "On some traffic equilibrium theory paradoxes," *Journal of Transportation Resource B*, vol. 18, pp. 101–110, 1984.

[16] M. J. Smith, "In a road network, increasing delay locally can reduce delay globally," *Journal of Transportation Resource*, vol. 12, pp. 419–422, 1978.

[17] E. Altman, R. E. Azouzi, and O. Pourtallier, "Avoiding paradoxes in multi-agent competitive routing," *Computer Networks*, vol. 43, no. 2, pp. 133–146, 2003.

[18] N. G. Bean, F. P. Kelly, and P. G. Taylor, "Braess' paradox in a loss network," Dec. 20 1995.

[19] B. Calvert, W. Solomon, and I. Ziedins, "Braess's paradox in a queuing network with state dependent routing," *Journal of Applied Probability*, vol. 34, pp. 134–154, 1997.

[20] P. Jacquet, "Geometry of information propagation in massively dense ad hoc networks," in *MobiHoc*, 2004, pp. 157–162.

[21] M. Kalantari and M. Shayman, "Energy efficient routing in sensor networks," in *Proc. of Conference on Information Sciences and Systems.* Princeton University, March 2004.

[22] ——, "Routing in wireless ad hoc networks by analog to electrostatics theory," in *Proceedings of IEEE International Communications Conference ICC'04*, Paris, France, June 2004.

[23] S. Toumpis, "Optimal design and operation of massively dense wireless networks: or how to solve 21st century problems using 19th century mathematics," in *Interperf '06: Proceedings from the 2006 workshop on Interdisciplinary systems approach in performance evaluation and design of computer & communications sytems.* New York, NY, USA: ACM, 2006, p. 7.

[24] S. Toumpis and L. Tassiulas, "Optimal deployment of large wireless sensor networks," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2935–2953, 2006.

[25] G. A. Gupta and S. Toumpis, "Optimal placement of nodes in large sensor networks under a general physical layer model," in *IEEE Secon*, Santa Clara, CA, September 2005.

[26] S. Toumpis, "Mother nature knows best: A survey of recent results on wireless networks based on analogies with physics," *Computer Networks*, vol. 52, no. 2, pp. 360–383, 2008.

[27] E. Hyytiä and J. Virtamo, "On load balancing in a dense wireless multihop network," in *NGI 2006, 2nd Conference on Next Generation Internet Design and Engineering*, València, Spain, Apr. 2006, pp. 72–79.

[28] M. Beckmann, "A continuum model of transportation," *Econometrica*, vol. 20, no. 4, pp. 643–660, 1952.

[29] S. C. Dafermos, "Continuum modeling of transportation networks," *Transportation Research*, vol. 14B, pp. 295–301, 1980.

[30] P. Daniele and A. Maugeri, "Variational inequalities and discrete and continuum models of network equilibrium protocols," *Mathematical and Computer Modelling*, vol. 35, pp. 689–708, 2002.

[31] "Two-dimensional continuum modeling approach to transportation problems," *Journal of Transportation Systems Engineering and Information Technology*, vol. 6, no. 6, pp. 53 – 68, 2006.

[32] G. Idone, "Variational inequalities and applications to a continuum model of transportation network with capacity constraints," *J. of Global Optimization*, vol. 28, no. 1, pp. 45–53, 2004.

[33] S. Wong, Y. Du, L. Sun, and B. Loo, "Sensitivity analysis for a continuum traffic equilibrium problem."

[34] W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, *Combinatorial Optimization*. New York: John Wiley, 1997.

[35] M. Bardi and I. C. Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Boston, Massachussets: Birkhäuser, 1997.

[36] W. Fleming and H. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed. Springer, 2006.

[37] S. C. Wong, "Multi-commodity traffic assignment by continuum approximation of network flow with variable demand," *Transportation Research Part B: Methodological*, vol. 32, no. 8, pp. 567 – 581, 1998.

[38] S. C. Wong, C. K. Lee, and C. O. Tong, "Finite element solution for the continuum traffic equilibrium problems," *International Journal for Numerical Methods in Engineering*, vol. 43, no. 7, pp. 1253–1273, 1998.

[39] H. Yang and H.-J. Huang, "The multi-class, multi-criteria traffic network equilibrium and systems optimum problem," *Transportation Research Part B: Methodological*, vol. 38, no. 1, pp. 1 – 15, 2004.

[40] R. J. Vaughan, *Urban Spatial Traffic Patterns*. London: Pion, 1987.

[41] Energetics, "Itp sensors and automation: Industrial wireless technology for the 21st century." [Online]. Available: http://www.energetics.com/pdfs/technologie_processes/wireless.pdf

[42] L. Popa, A. Rostamizadeh, R. Karp, C. Papadimitriou, and I. Stoica, "Balancing traffic load in wireless networks with curveball routing," in *MobiHoc*, E. Kranakis, E. M. Belding, and E. Modiano, Eds. ACM, 2007, pp. 170–179. [Online]. Available: http://doi.acm.org/10.1145/1288107.1288131

[43] E. Hyytiä and J. T. Virtamo, "On the optimality of field-line routing in massively dense wireless multi-hop networks," *Perform. Eval.*, vol. 66, no. 3-5, pp. 158–172, 2009.

[44] S. Durocher, E. Kranakis, D. Krizanc, and L. Narayanan, "Balancing traffic load using one-turn rectilinear routing," in *TAMC*, ser. Lecture Notes in Computer Science, M. Agrawal, D.-Z. Du, Z. Duan, and A. Li, Eds., vol. 4978. Springer, 2008, pp. 467–478. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-79228-4_41

[45] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, no. 2, pp. 388–404.

[46] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1009–1018, 2007.

[47] F. Baccelli, B. Blaszczyszyn, and P. Mühlethaler, "An aloha protocol for multihop mobile wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 421–436, 2006.

[48] S. Toumpis and L. Tassiulas, "Packetostatics: deployment of massively dense sensor networks as an electrostatics problem," in *INFOCOM*, 2005, pp. 2290–2301.

[49] J. A. Silvester and L. Kleinrock, "On the capacity of multi-hop slotted ALOHA networks with regular structure," *IEEE trans. on commun.*, vol. COM-31, pp. 974–982, 1983.

[50] V. Ramaiyan, A. Kumar, and E. Altman, "Jointly optimal power control and hop length for a single cell, dense, ad hoc wireless network," in *Proc. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, WiOpt'07*, Limassol, Cyprus, April 2007.

[51] V. Rodoplu and T. H. Meng, "Bits-per-joule capacity of energy-limited wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 857–865, 2007.

[52] R. Catanuto, S. Toumpis, and G. Morabito, "On asymptotically optimal routing in large wireless networks and geometrical optics analogy," *Computer Networks*, vol. 53, no. 11, pp. 1939–1955, 2009.

[53] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multi-user communication networks," in *INFOCOM*, 1993, pp. 964–971.

[54] P. Gupta and P. R. Kumar, "A system and traffic dependent adaptive routing algorithm for ad hoc networks," in *Proc. of the 36th IEEE Conference on Decision and Control*, San Diego, CA, USA, December 1997, pp. 2375–2380.

[55] H. Weinberger, *A First Course in Partial Differential Equations with Complex Variables and Transform Methods*. New York: Dover Publications Inc., 1995, corrected reprint of the 1965 original.

[56] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton University Bulletin*, pp. 49–52, 1902.

[57] L. C. Evans, *Partial Differentiali Equations*. Graduate Studies in Mathematics, American Mathematical Society, vol. 19.

[58] A. Haurie and P. Marcotte, "On the relationship between nash cournot and wardrop equilibria," *Networks*, vol. 15, no. 3, pp. 295–308, 1985.

[59] M. Patriksson, "The traffic assignment problem: Models and methods," *VSP BV*.

[60] V. Girault., *Approximations variationnelles des E.D.P.* Cours DEA, 2005-2006.

[61] E. Altman, P. Bernhard, and A. Silva, "The mathematics of routing in massively dense ad-hoc networks," in *ADHOC-NOW '08: Proceedings of the 7th international conference on Ad-hoc, Mobile and Wireless Networks*, 2008, pp. 122–134.

[62] H. Brezis, *Analyse fonctionelle*, ser. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Paris: Masson, 1983.

[63] BBC-News, "Mobile system promises free calls." [Online]. Available: http://news.bbc.co.uk/2/hi/technology/6987784.stm,March30,2010

[64] "More information about the one laptop per child program." [Online]. Available: http://olpc.com/andhttp://laptop.org/en/,March30,2010.

[65] R. J. D. Perna and P. L. Lions, "Ordinary differential equations, transport theory and sobolev spaces," *Inventiones mathematicae*, vol. 98, pp. 511–547, 1989.

[66] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002. [Online]. Available: http://dx.doi.org/10.1002/wcm.72

[67] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 1991, vol. 113.

[68] F. Santambrogio, “Models and applications of optimal transport theory,” in *Lecture Notes at Summer School 2009.* [Online]. Available: http://www-fourier.ujf-grenoble.fr/IMG/pdf/Santambrogio2.pdf

[69] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423 & 623–656, 1948.

[70] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” Jan. 29 1998. [Online]. Available: http://www.bell-labs.com/project/blast/wpc-v6n3.pdf

[71] I. E. Telatar, “Capacity of multi-antenna gaussian channels,” Mar. 14 2000. [Online]. Available: http://mars.bell-labs.com/cm/ms/what/mars/papers/proof/proof.ps.gz

[72] M. Debbah and R. R. Müller, “MIMO channel modeling and the principle of maximum entropy,” *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1667–1690, 2005. [Online]. Available: http://dx.doi.org/10.1109/TIT.2005.846388

[73] L. Schumacher, K. Pedersen, and P. E. Mogensen, “From antenna spacings to theoretical capacities - guidelines for simulating mimo systems,” in *PIMRC*, 2002.

[74] T. S. Pollock, T. D. Abhayapala, and R. A. Kennedy, “Antenna saturation effects on dense array mimo capacity,” Oct. 24 2003.

[75] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna gaussian broadcast channel,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[76] A. M. Tulino and S. Verdú, “Random matrix theory and wireless communications,” *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[77] E. Wigner, “Random matrices in physics,” *SIAM Review*, vol. 9, no. 1, January, 1967.

[78] J. W. Silverstein and Z. D. Bai, “On the empirical distribution of eigenvalues of a class of large dimensional random matrices,” Jan. 13 1995. [Online]. Available: http://citeseer.ist.psu.edu/265596.html;http://www4.ncsu.edu/~jack/mandp.pdf

[79] V. L. Girko, *Theory of random determinants. Transl. from the Russian.* Mathematics and Its Applications (Soviet Series), 45, 1988.

[80] P. Dent, G. E. Bottomley, and T. Croft, “Jakes fading model revisited,” *Electronic Letters*, vol. 29, no. 13, pp. 1162–1163, June 1993.

[81] Jaynes, *Probability Theory: The Logic of Science.* Cambridge University Press, 2003.

[82] M. Guillaud, M. Debbah, and A. L. Moustakas, “Maximum entropy MIMO wireless channel models,” *CoRR*, vol. abs/cs/0612101, 2006, informal publication. [Online]. Available: http://arxiv.org/abs/cs/0612101,February,2010.

[83] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2491–2509, 2005.

[84] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, 2005. [Online]. Available: http://dx.doi.org/10.1561/0100000006

[85] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge University Press, 2005.

[86] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[87] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TCOMM.2004.840638

[88] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," Jan. 1998. [Online]. Available: http://ProjectEuclid.org/getRecord?id=euclid.aop/1022855421

[89] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[90] G. Buttazzo and F. Santambrogio, "A model for the optimal planning of an urban area," *SIAM J. Math. Anal.*, vol. 37, no. 2, pp. 514–530 (electronic), 2005. [Online]. Available: http://dx.doi.org/10.1137/S0036141003438313

[91] G. Carlier and I. Ekeland, "The structure of cities," *J. Global Optim.*, vol. 29, no. 4, pp. 371–376, 2004. [Online]. Available: http://dx.doi.org/10.1023/B:JOGO.0000047909.02031.ab

[92] ——, "Equilibrium structure of a bidimensional asymmetric city," *Nonlinear Anal. Real World Appl.*, vol. 8, no. 3, pp. 725–748, 2007. [Online]. Available: http://dx.doi.org/10.1016/j.nonrwa.2006.02.008

[93] G. Carlier, C. Jimenez, and F. Santambrogio, "Optimal transportation with traffic congestion and Wardrop equilibria," *SIAM J. Control Optim.*, vol. 47, no. 3, pp. 1330–1350, 2008. [Online]. Available: http://dx.doi.org/10.1137/060672832

[94] Y. Yu, B. Danila, J. A. Marsh, and K. E. Bassler, "Optimal transport on wireless networks," Mar. 28 2007, comment: 5 pages, 4 figures. [Online]. Available: http://arxiv.org/abs/physics/0703261

[95] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks, Volume I - Theory*, ser. Foundations and Trends in Networking Vol. 3: No 3-4, pp 249-449, F. Baccelli and B. Blaszczyszyn, Eds. NoW Publishers, 2009, vol. 1, Stochastic Geometry and Wireless Networks, Volume II - Applications; see http://hal.inria.fr/inria-00403040. [Online]. Available: http://dx.doi.org/10.1561/1300000006http://hal.inria.fr/inria-00403039/en/

[96] T. S. Rappaport, *Wireless Communication Principles and Practice.* Prentice Hall, 1996.

[97] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley & Sons, N.Y., 1991.

[98] G. S. Kasbekar, E. Altman, and S. Sarkar, "A hierarchical spatial game over licenced resources," in *GameNets'09: Proceedings of the First ICST international conference on Game Theory for Networks.* Piscataway, NJ, USA: IEEE Press, 2009, pp. 70–79.

[99] E. Altman, A. Kumar, C. K. Singh, and R. Sundaresan, "Spatial SINR games combining base station placement and mobile association," in *INFOCOM.* IEEE, 2009, pp. 1629–1637. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2009.5062081

[100] G. Monge, "Mémoire sur la théorie des déblais et des remblais," *Histoire de l'Académie Royale des Sciences*, 1871.

[101] Y. Brenier, "Décomposition polaire et réarrangement monotone des champs de vecteurs," *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 305, no. 19, pp. 805–808, 1987.

[102] C. Villani, *Optimal transport*, ser. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Berlin: Springer-Verlag, 2009, vol. 338, old and new. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-71050-9